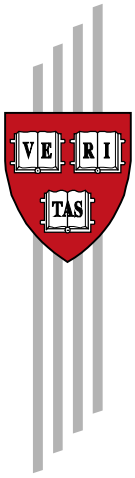
Bridging the short-term and long-term dynamics of economic structural change

James McNerney, Yang Li, Andres Gomez-Lievano,
and Frank Neffke

CID Research Fellow and Graduate Student
Working Paper No. 133
October 2021

© Copyright 2021 McNerney, James; Li, Yang;
Gomez-Lievano, Andres; Neffke, Frank; and the President and
Fellows of Harvard College



Working Papers

Center for International Development
at Harvard University

Bridging the short-term and long-term dynamics of economic structural change

James McNerney,[1,*] Yang Li,[1] Andrés Gómez-Liévano,[1] and Frank Neffke[1,2]

[1] *Center for International Development, Kennedy School of Government, Harvard University, Cambridge MA 02139, USA*

[2] *Complexity Science Hub Vienna, Vienna, Austria*

In the short-term, economies shift preferentially into new activities that are related to ones they currently do. Such a tendency should have implications for the nature of an economy's long-term development as well. We explore these implications using a dynamical network model of an economy's movement into new activities. First, we theoretically derive a pair of coordinates that summarize long-term structural change. One coordinate captures overall ability across activities, the other captures an economy's composition. Second, we show empirically how these two measures intuitively summarize a variety of facts of long-term economic development. Third, we observe that our measures resemble complexity metrics, though our route to these metrics differs significantly from previous ones. In total, our framework represents a dynamical approach that bridges short- and long-term descriptions of structural change, and suggests how different branches of economic complexity analysis could potentially fit together in one framework.

The prosperity of an economy is closely tied to the economic activities it can develop[1]. Whereas places like Silicon Valley, the city of London, and the country of Japan pursue diverse and profitable activities, other places struggle to shift out of a narrow range of activities with low economic returns. Working to understand why, recent research has emphasized that economic diversification typically entails shifts into activities that are related to ones that are already in place. This tendency has been corroborated in research on the growth of industry clusters[2,3], and on related diversification by individuals[4,5], firms[6–8], regions[9–13], and countries[14–16]. Recently these phenomena have been referred to collectively as the Principle of Relatedness[17].

We interpret the Principle of Relatedness (PoR) as a stylized empirical regularity of short-term structural change. We then use it to explore implications for how economies would be expected to develop over longer periods of time. Our strategy is as follows. We define a dynamical model that implements the PoR, using as a guide the mathematical structure of empirical regression setups that test this principle in the literature. We then follow a standard recipe: We exploit the linear structure of these models, and analyze our model using eigenmode decomposition. As in any dynamical system, each dynamical mode in our diversification model has a characteristic time scale. A well-appreciated point is that system-changes over long times are dominated by the modes characterized by the longest time scales. Applying this general principle of dynamical systems to our PoR-based model, we focus on the two slowest modes of activity transitions expected under our theory. Projecting an economy onto these eigenmodes then yields coordinates, which describe where the economy is situated along these axes of change. These coordinates summarize the main features of an economy's activity basket in a way that is especially suited to describing how this basket changes over long periods of time.

Interestingly, the modes of change that we focus on have simple interpretations. One corresponds to an *average* increase in an economy's ability across many different activities. We show empirically that it is associated with increases in the economy's diversity. The other mode of change corresponds to a compositional shift in the activity basket: It quantifies how the *relative* importance of activities changes vis-à-vis one another.

To gain further understanding, we apply this framework in the empirical setting of country export baskets. In particular, we summarize 50 years' worth of high-dimensional change in the global export economy in a 2D plane. The axes of this plane correspond to the two slowest eigenmodes, and coordinate pairs in this plane depict the export economy of individual countries. One coordinate tracks the general ability level of a country across activities, the other tracks information about activity composition. We find that the dominant movements we empirically observe in this plane are intuitive and consistent with classic observations about economic development. For instance, the second eigenmode captures a shift out of agriculture and into manufacturing. Consequently, simultaneous movements along the first and second eigenmode correspond to a well-documented pattern of development in which countries simultaneously diversify into new products while shifting into manufactured goods (e.g.[18,19]).

Surprisingly, these coordinates turn out to be closely related to *complexity metrics* that have been proposed in recent years[20–26]. In particular, the compositional mode of change described above is closely related to measures known as the Product Complexity Index (PCI) and Economic Complexity Index (ECI)[20,27–29]. While our dynamical theory leads to quantities that closely resemble the PCI and ECI, the corresponding quantities in our dynamical theory, PCI* and ECI*, have rather different mathematical and conceptual underpinnings. First, the ECI and PCI have been motivated by discussions of how information about the productive knowledge, sophistication, or diversity of capabilities that different products require and different economies possess can be inferred

* james_mcnerney@hks.harvard.edu

by observing what economic activities are performed where. However, in our framework, an eigenvector whose elements resemble the PCIs emerges for conceptually different reasons: It describes a pattern of activity shifts that takes economies an especially long time to execute. The associated coordinate, in turn, expresses how far a country has progressed in executing these shifts, and turns out to closely resemble the ECI in functional form and to strongly correlate with it numerically. Second, the conventional PCI and ECI can be seen as interdependent metrics[30] and their mathematical definitions are mirror images of one another. In our approach, in contrast, PCI* and ECI* have distinct characters: The PCI*s take on collective meaning as describing a pattern or direction of economic change, and the ECI* of any economy individually captures its progress in this direction. We discuss these differences further, as well as empirical relationships between our coordinates and complexity metrics more generally, including country Fitness[21], Production Ability[25], and GENEPY[26]. We observe that existing complexity metrics fall into two groups that correlate strongly with the two coordinates expected by our theory, with one set of quantities emphasizing the diversity of economic activities, and the other set of quantities emphasizing their composition.

I. BACKGROUND

A widespread finding about the geography of economic activity is that particular activities tend to coincide in the same places. These patterns are often intuitive; a city, region, or country that makes cloth garments, for example, is also likely to make other textiles (e.g. knitted garments). Such activities are said to be related and various measures of *relatedness* have been developed. For concreteness, we focus on relatedness between exported products as inferred from the co-occurrence of these products in the export baskets of countries. Studies typically begin by measuring activity in a product using the Balassa index of revealed comparative advantage (RCA)[31], $R_{cp} = \theta_{cp}/\theta_p$, where θ_{cp} is the share of country c's exports devoted to product p, and θ_p is the share of p in world exports. RCAs are often treated as measures of inferred ability in a product[32,33], and these quantities have also been used to characterize ability in activities besides exports (where the meanings of θ_{cp} and θ_p are adjusted appropriately). Here we follow suit though it is not an intrinsic requirement of our model to measure ability this way.

The relatedness, or proximity, of products p and p' is then computed by a measure of similarity between the p and p' columns of the matrix $R = [R_{cp}]$. Thus, p and p' are estimated to be highly related if high ability in p frequently co-occurs with high ability in p' across locations. For example, Hidalgo et al. (2007)[15] say that a country has significant ability in a product if its RCA in the product exceeds 1, and then compute the conditional probability $Q_{pp'}$ that country c has an RCA greater than 1 in good p given that it also satisfies this condition for p'. The proximity between products p and p' is then taken to be the lesser of the two conditional probabilities $Q_{pp'}$ and $Q_{p'p}$: $\Phi^M_{pp'} \equiv \min(Q_{pp'}, Q_{p'p})$. These proximities are dominated by small values and sparse in high values, motivating network representations of Φ^M (Fig. 1).

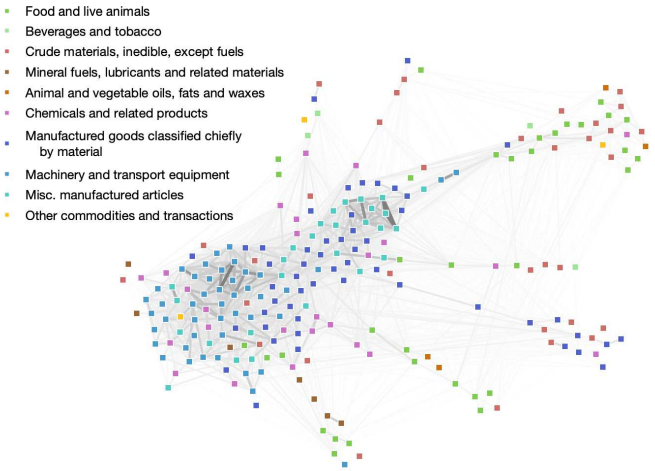


FIG. 1. Network representation of relatedness between exported products, computed from $\Phi^M_{pp'} \equiv \min(Q_{pp'}, Q_{p'p})$ (see text) using UN Comtrade data[36].

Crucially, changes in an economy's activities are predicted by the structure of proximities, such that economies shift preferentially into activities that are related to ones they do already. Studies show that the development of high ability in an exported product tends to be preceded by high ability in nearby products[15]. A typical regression setup[11,13,34,35] to explore this effect takes future ability in a product p to be a function of the *density* around it. This is defined as the average ability that an economy has in other products, with each product weighted by its proximity to p:

$$R_p(t+1) \sim b_1 R_p(t) + b_2 \sum_{p'} \frac{\Phi_{pp'}}{\sum_{p''} \Phi_{pp''}} R_{p'}(t). \qquad (1)$$

An alternate setup[16] contains two steps, with ability in product p first regressed on density in the same time period, establishing the existence of systematic correlations in which economic activities co-occur. The residuals from this regression are then used to forecast appearances of comparative advantage in future time periods. Combining two such regressions analytically will also lead to an expression of the form of Eq. (1). Empirical support for the density terms in these regressions has bolstered the idea that relatedness measures capture underlying similarities in activities that make some transitions easier to achieve than others, influencing the direction in which an economy develops.

II. RESULTS

A. PoR as a dynamical model

The PoR suggests that diversification to new activities can be described as a process of spreading on a network of activities. A given economy has various levels of ability in different activities, which are given by the elements of a vector $\mathbf{R}(t)$. We refer to this as the economy's activity basket. These abilities evolve as the economy improves or shifts into new parts of the network. Using Eq. (1) as a guide, a simple model of this spreading process is

$$\dot{\mathbf{R}}(t) = \gamma(t)\mathbf{R}(t) - \frac{1}{\tau}L_\Phi \mathbf{R}(t), \tag{2}$$

where $\gamma(t)$ is an arbitrary growth rate, L_Φ is a graph Laplacian, and τ is a time scale. The two terms in Eq. (2) captured two kinds of changes: An economy's abilities can (generally) rise or fall as a whole, and they can also shift to weight activities differently, changing which ones receive the most emphasis.

The Laplacian term captures an implicit choice of empirical specifications for the PoR as instances of consensus dynamics. The Laplacian L_Φ captures the structure of the network and governs shifts between activities, taking the form $L_\Phi = I - \tilde{\Phi}$ where $\tilde{\Phi}$ is a proximity matrix whose rows have been normalized to sum to 1, $\tilde{\Phi}_{pp'} = \Phi_{pp'} / \sum_{p''} \Phi_{pp''}$. With this assumption, a discrete-time approximation of Eq. (2) in index form reads

$$\frac{R_p(t+\Delta t) - R_p(t)}{\Delta t} \approx \gamma(t)R_p(t) + \frac{1}{\tau}\left(\sum_{p'} \tilde{\Phi}_{pp'}R_{p'}(t) - R_p(t)\right). \tag{3}$$

In this form our dynamical model is more easily compared with the regression setups that test for the PoR in the literature. Ability in activity p changes between the periods t and $t+\Delta t$ in a way that depends on ability in nearby products. When the average ability in activities around p ($\sum_{p'} \tilde{\Phi}_{pp'}R_{p'}(t)$) exceeds that in p itself ($R_p(t)$), ability in p rises. In addition, abilities can rise or fall as a whole, and this is captured by the term $\gamma(t)R_p(t)$.

Integrating Eq. (2) over time leads to the expression

$$\mathbf{R}(t) = A(t)P(t)\mathbf{R}(0), \tag{4}$$

where the dynamics captured by the two terms in Eq. (2) are now captured by the scalar prefactor $A(t)$ and the matrix $P(t)$. The prefactor $A(t) \equiv e^{\int_0^t \gamma(s)\, ds}$ is a shift factor that captures the accumulated growth in overall ability up to time t, scaling the $\mathbf{R}$ vector up or down as a whole. The matrix $P(t) = e^{-L_\Phi(t/\tau)}$ is a stochastic matrix that transforms $\mathbf{R}$, changing the relative emphasis on different activities over time.

Implementing this model involves some practical considerations. First, RCAs are commonly trans formed non-linearly to weaken the influence of extreme values[16,19,37,38]. When we use RCAs as a measure of ability, we will do so as well, but note that this does not change the model in any fundamental way. Second, we need to choose how to operationalize the network of relatedness $\Phi_{pp'}$ between activities. Studies have used a variety of alternative measures of proximity, such as ones based on correlations[16]. Here we consider several options, finding similar outcomes for our results, as we discuss later.

B. Eigenmode decomposition

Dynamical models are frequently analyzed in terms of their dynamical modes, often called eigenmodes in linear models. Applying eigenmode decomposition to the model above sets up coordinates to describe changes in activity baskets over long time scales that will form the basis for the empirical analysis later on. Let $\mathbf{v}_\mu$ be the μth right eigenvector of L_Φ and let κ_μ be its eigenvalue. Ordering eigenvalues from least to greatest, the first right eigenvector has $\kappa_1 = 0$ and is a uniform vector of real positive numbers. It can be taken to be a vectors of 1s, $\mathbf{v}_1 = \mathbf{1}$. Higher-order eigenvectors ($\mathbf{v}_2$ and up) have $\kappa_\mu > 0$, and contain a mix of elements with positive and negative real parts.[1]

The right eigenvectors of L_Φ form a basis for the vector space of ability vectors $\mathbf{R}(t)$, and as a result, one can write any such vector as a linear combination of eigenvectors,

$$\mathbf{R}(t) = \sum_\mu c_\mu(t)\mathbf{v}_\mu. \tag{5}$$

Letting $\mathbf{w}_\mu$ be the μth *left* eigenvector of L_Φ, the coefficients $c_\mu(t)$ above may be computed by exploiting the biorthogonality of L_Φ's left and right eigenvectors, $\mathbf{w}_\mu^\dagger \mathbf{v}_\nu = \delta_{\mu\nu}$:

$$c_\mu(t) = \mathbf{w}_\mu^\dagger \mathbf{R}(t). \tag{6}$$

Proceeding with the usual steps of eigenmode decomposition, we decompose $\mathbf{R}(t)$ to separate different modes of change according to their time scales. Plugging Eq. (5) into Eq. (4) gives $\mathbf{R}(t) = A(t)\sum_\mu c_\mu(0)e^{-\kappa_\mu t/\tau}\mathbf{v}_\mu$, where the first term contains the first eigenvector $\mathbf{v}_1 = \mathbf{1}$ and corresponds to the fixed point of the simpler model $\mathbf{R}(t) = P(t)\mathbf{R}(0)$ (i.e. the model with the factor $A(t)$ fixed at 1). Separating the first term from other terms of the sum, we have

$$\mathbf{R}(t) = A(t)\left(c_1(0)\mathbf{1} + \sum_{\mu\geq 2} c_\mu(0)e^{-t/\tau_\mu}\mathbf{v}_\mu\right), \tag{7}$$

[1] Some eigenvectors may contain imaginary parts, but this poses no special difficulties in interpreting modes. These eigenvectors occur in complex conjugate pairs and can be summed in pairs so that imaginary parts cancel (see Supplementary Material). In practice though none of the eigenvectors we are interested in here contain imaginary components.

where we defined $\tau_\mu \equiv \tau/\kappa_\mu$.

Eq. (7) describes the following behavior. First, neglecting the effect of the shift factor $A(t)$, the activity basket of a region eventually converges to a uniform vector $c_1(0)\mathbf{1}$. In this state, the region has equal ability in all products. On its way to this state, the basket shows higher ability in some products and lower ability in others. Each eigenvector $\mathbf{v}_\mu$ describes a specific pattern of deviations from the long-term steady state situation in which all activities are equally important. Since each eigenvector $\mathbf{v}_\mu$ for $\mu \geq 2$ satisfies $\mathbf{w}_1^\dagger \mathbf{v}_\mu = 0$, where $\mathbf{w}_1$ has only positive elements, each such eigenvector necessarily has some elements that are positive and others that are negative. Each of these patterns of deviation decays with time at the rate set by τ_μ, the characteristic time scale of the μth eigenmode. If the pre-factor $A(t)$ were fixed and equal to 1, then a region would be destined to have ability $c_1(0)$ in every product. However, letting $A(t)$ change over time, the final level of ability in products can be arbitrarily high or low.

C. A coordinate system for activity baskets

To probe the long-term implications of the PoR for structural change, we now focus on the first two eigenvectors of the expansion, i.e. the two eigenvectors associated with dynamics on the longest time scales. Absorbing the coefficient $c_1(0)$ into $A(t)$, and defining the coefficient $b(t) \equiv c_2(0)e^{-t/\tau_2}$, the longest-lived dynamics of $\mathbf{R}(t)$ are described by the two leading terms of the eigenexpansion:

$$\mathbf{R}(t) \sim A(t)\mathbf{1} + A(t)b(t)\mathbf{v}_2. \tag{8}$$

Here, the coefficients $A(t)$ and $b(t)$ have simple interpretations. A shift in $A(t)$ corresponds to a region realizing a uniform change in its abilities across activities. A shift in $b(t)$ corresponds to a compositional shift. Some activities rise in ability and others fall, as determined by the signs and magnitudes of entries in $\mathbf{v}_2$. Together, these coordinates situate the activity basket $\mathbf{R}(t)$ in a 2D space.

To further interpret the $A(t)$ coordinate, let $\boldsymbol{\pi} = \mathbf{w}_1$ be the first left eigenvector of L_Φ, normalized so that its elements sum to 1, and note that $\bar{R}(t) = \boldsymbol{\pi}^T\mathbf{R}(t)$ defines a weighted average of the abilities of a region. Again exploiting the biorthogonality of left and right eigenvectors, it can be shown (see Supplementary Information) that $A(t)$ is equal to this average:

$$A(t) = \bar{R}(t) \equiv \boldsymbol{\pi}^T\mathbf{R}(t). \tag{9}$$

For this reason, we refer to $A(t)$ as the average ability coordinate.

To further interpret the $b(t)$ coordinate, note that the RCA vector conveys two kinds of information about an economy's activities: It conveys information about the relative mix or composition of activities, and information about ability levels. Arguably, a more obvious way to study an economy's activity basket would be to examine its shares of output in each activity, but unlike shares, RCAs are not constrained to sum to 1. An economy can have large RCAs in many activities, signaling that it is good at many things, or have many small RCAs. Factoring out an economy's average ability $A(t)$ from the vector $\mathbf{R}(t)$ gives a normalized version of this vector, $\mathbf{r}(t)$, that only characterizes the composition of activities: $\mathbf{r}(t) \equiv \mathbf{R}(t)/A(t)$. Dividing any of Eqs. (4), (7), or (8) by $A(t)$ leads to a nearly identical equation for $\mathbf{r}(t)$ where the factor $A(t)$ is absent. In particular, dividing Eq. (8) by $A(t)$ gives $\mathbf{r}(t) \sim \mathbf{1} + b(t)\mathbf{v}_2$, showing that $b(t)$ characterizes the deviation of the compositional vector $\mathbf{r}(t)$ from a uniform vector.

D. Evolution of country exports baskets

To explore how these coordinates behave in a practical setting, we apply them to the evolution of countries' export baskets in the world economy, using UN Comtrade data[36] from 249 regions over the period 1962 - 2018. This exercise summarizes country development by two simple statistics.

RCAs are commonly transformed to mitigate the influence of extreme values, and here we transform the heavy-tailed RCAs using the function $g(R) \sim \log(1 + R/R_0)$ (see Supplementary Information). We tune the parameter R_0 such that it maximizes the amount of variance that the b coordinate can explain across time and countries. We then calculate for each country c and time t the RCA vector $\mathbf{R}_c(t)$ on this transformed scale and use these vectors to compute the A and b coordinates.

Figs. 2a-b depict country income and export diversity as functions of A and b. For reasons that will become clear in the next section, we also refer to b as ECI* in this graph. Plotting income and diversity this way allows us to ask how each of these variables vary as a function of one coordinate while holding the other fixed.

Income increases with b, diversity with A. A higher value of the compositional coordinate b for a country's export basket is associated with significantly higher GDP per capita (Fig. 2a). For example, holding A fixed at 0.5, an increase in b from -0.5 to +0.5 is associated with an increase from about \$1500 to \$30,000 (2010 US dollars PPP). In contrast, the association between a country's average ability A and its income when holding the value of b fixed is weak at best.

The situation is roughly opposite when we examine countries' economic diversity in these coordinates (Fig. 2b). The diversity of a country's economic activity has been quantified with a number of measures, such as the Gini coefficient and Herfindahl-Hirschmann index (e.g.[18]), or the count of products d_c in which a region c has an RCA greater than 1. Higher values of the average ability coordinate A are closely associated with greater diversity in a country's export basket. Holding b fixed at 0, an increase in A from 0.1 to 0.9 is associated with a more-than-10 fold rise in number of products with RCA above 1. In contrast, when holding A fixed, a compositional

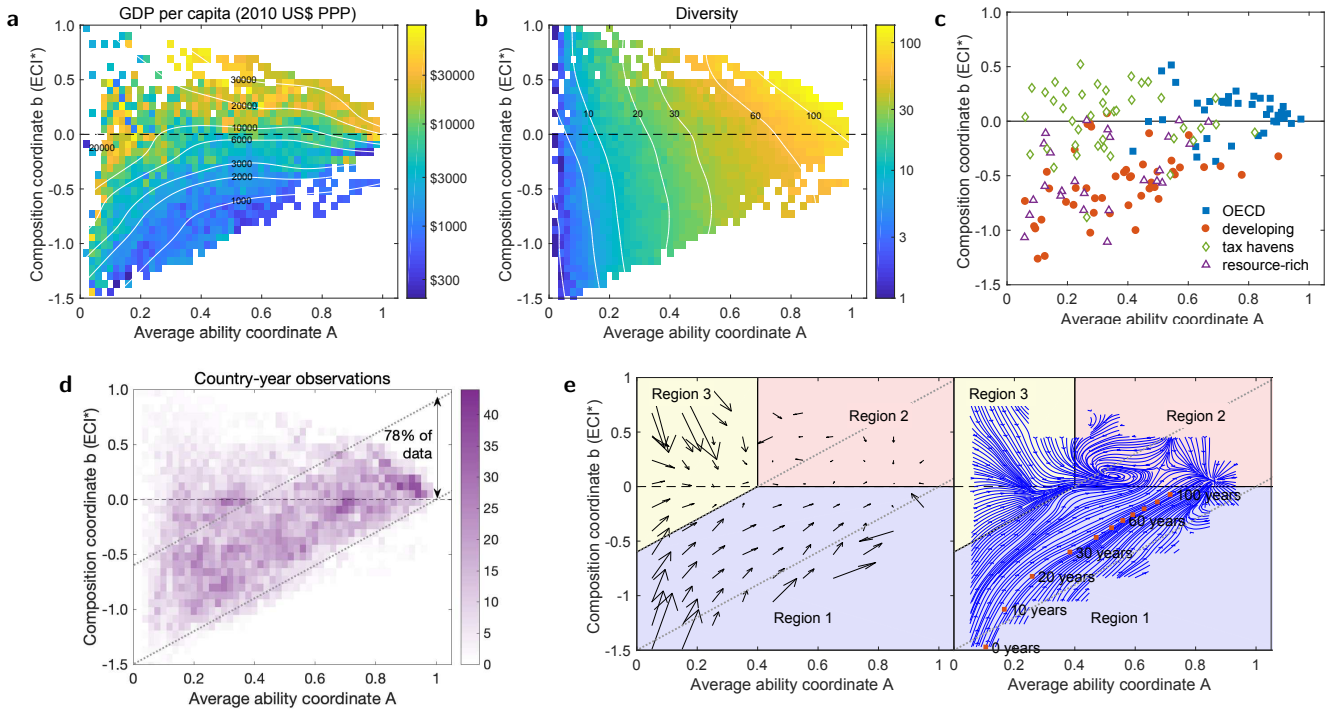


FIG. 2. Country export baskets plotted according to average ability (A coordinate) and composition (b coordinate), based on Comtrade data[36] for 249 countries and regions over the period 1962 - 2018 (11,544 observations). **a-b** We form bins in the 2D plane to collect country-year observations, and compute average GDP per capita and average diversity d within each bin. White lines are contours of a LOWESS-smoothed surface fit to bin averages. **c** Positions of export baskets in 2018 of four categories of economies. 'Developing' refers to economies on the UN's least developed economies list[39], 'tax havens' to non-OECD and non-developing countries appearing in Hines (2010)[40], and 'resource-rich' to economies not elsewhere categorized whose natural resource rent[41] exceeded 5% of GDP in 2018. **d** Number of country-year observations in each bin. **e** Quiver and streamline plots for the average directional change over 20 years for countries beginning at different starting points in the plane. Red dots in the right panel follow one streamline and roughly convey the average pace of movement, though no streamline should be taken as a typical country trajectory; individual countries display wide variation in both direction and speed of change.

shift towards higher levels of b is only weakly associated with higher export diversity.

Different types of countries inhabit different regions of this coordinate system (Fig. 2c). The industrialized countries of the OECD are predominantly located in the right of the plot with relatively high values of both A and b, corresponding to diverse, high-income, developed economies. In the lower left, we find a set of undiversified developing economies as defined by the UN's least developed economies classification. In contrast, in the upper left, we see countries that are also undiversified, yet often have high-income. Many of these are often considered tax havens[40] or resource-rich economies.

Note that, although countries often diversify as they rise in income[18], the observations above do not directly associate higher income with higher diversity of products exported. Rather, they suggest that what matters more for a country's GDP per capita is the composition of exports[1]. We next explore the dynamics of countries in this space and probe this distinction further.

Country development. Export baskets occupy a triangular region of the A-b phase plane. A particularly densely occupied portion of this plane is a diagonal band that stretches from low A and low b, to high A and above-average b (Fig. 2d). We start by focusing on the movement of countries whose export baskets lie in this band. Countries show diverse trajectories (see Supplementary Information). To nevertheless characterize broad tendencies over long periods, we group country-year observations into bins in the A-b plane. Next, we examine the average direction and speed of movement over the next 20 years of observations that begin within a given bin. For expositional convenience, we summarize our results by referring to three regions of the plane, labelled Regions 1, 2, and 3 (Fig. 2e).

Countries beginning in the lower part of the main diagonal band – i.e. in Region 1 – trend over decadal time scales in a lower-left-to-upper-right direction. Given how the coordinates A and b are defined, a simultaneous increase in both directions corresponds to an export basket that simultaneously realizes two kinds of changes: (1) a general improvement in ability across products; (2) a shift in composition towards products with positive values in the vector $\mathbf{v}_2$. These two components of movement are associated with different effects. The shift in the composition of exports captured by increasing b is (empirically) most directly associated with increased GDP per capita (Fig. 2a). In contrast, the increase in A is more associated with a rise in export diversity (Fig. 2b).

These patterns are consistent with the idea that higher wealth is associated with exporting particular products, rather than diversification per se. Nevertheless, in practice, countries in the lower part of the main diagonal band that succeed in reaching such products tend do so while simultaneously diversifying into a broad range of goods, including ones not associated with higher wealth.

What kinds of products are involved in the vertical shifts that increase b? Manufacturing products very often have positive elements in $\mathbf{v}_2$, and agricultural products very often have negative elements. As a result, countries in Region 1 that traverse the length of the diagonal band see a broad, long-lived shift in their export baskets away from agricultural products and towards manufactured products. This shift is consistent with the long-observed tendency for economies to move from agriculture to manufacturing (and then on to services, a move that largely eludes trade statistics) as they develop[42,43].

If a country traverses the length of the diagonal band in Region 1 it will arrive in Region 2, where countries tend to be high-income, diverse, developed economies. Within this region, the average speed of movement is much lower than in Region 1. This in part owes to the fact that export baskets here evolve in a greater variety of directions, with an average directional change near to zero. Broadly though, countries in Region 2 (i) tend to sustain a high b, somewhat above 0, and (ii) move within a range of relatively high A values. On reaching Region 2, a number of countries move left toward lower values of A. This transformation path is consistent with a phenomenon in which countries see a fall in diversity in late stages of development[18,44].

Finally, countries in Region 3 tend to have high income and low diversity. Many of these countries are abundant in natural resources (particularly oil) or function as tax havens. In this region, countries tend to move quickly toward lower levels of b, converging near zero. We are not aware of any prior observations that this movement corresponds to, though apparently it is difficult for these countries to sustain a high degree of specialization in products that load positively on the b coordinate for long periods of time.

Taken together, these observations show an end-to-end agreement in our framework between the short-term and long-term in economic development. The Principle of Relatedness describes structural change on short time-scales, and fine-grained levels of sectoral resolution. When we analyze network models that operationalize this principle, using the standard tools of eigenmode decomposition, we arrive at coordinates that should capture activity changes over long time-scales, and at higher levels of aggregation. As one would hope, observed movements of export baskets in these coordinates do indeed reveal long-term and coarse-grained descriptions of structural change that are consistent with well-documented stylized facts.

E. Complexity metrics

We have discussed that the second eigenvector $\mathbf{v}_2$ of the diversification model Eq. (2) captures the mode of compositional change in activities that will take an economy the longest time to execute. The signs and magnitudes of the elements of $\mathbf{v}_2$ convey a pattern of structural transformation that describes how countries shift in relative terms out of some activities (associated with the negative elements of $\mathbf{v}_2$), while entering others (the positive elements). Progress of a country along this direction of change is captured by $b(t)$. Surprisingly, the quantities $\mathbf{v}_2$ and $b(t)$ that describe this shift from some activities to others turn out to be closely related to the complexity metrics known as the PCI and ECI.

The PCI and ECI. The PCI and the ECI were proposed by Hidalgo and Hausmann (2009)[20], and can be computed with an eigenvector computation as noted by Caldarelli et al. (2012)[28,29]. To briefly summarize the calculation of these indices, let n_c be the number of countries and n_p the number of products. Furthermore, let R be the $n_c \times n_p$ matrix R of RCAs, and let M be a binarized version of this matrix, with elements $M_{cp} = 1$ wherever $R_{cp} \geq 1$ and 0 otherwise. M may be viewed as an adjacency matrix for a bipartite network connecting countries to products that they export at a significant ability level. Following Hidalgo and Hausmann (2009)[20], we define a row sum of M as a country's export *diversity*, $d_c \equiv \sum_p M_{cp}$. Similarly, we define a column sum of M as a product's *ubiquity*, $u_p \equiv \sum_c M_{cp}$.

Let D and U be the diagonal matrices formed from countries' diversities and products' ubiquities, respectively. One may multiply $D^{-1}M$ and $U^{-1}M^T$, in different orders, to define two square matrices that eliminate either the country nodes or the product nodes from the bipartite network. Multiplying these matrices in one order collapses the country dimension, leading to the $n_p \times n_p$ row-normalized stochastic matrix $\mathcal{P} \equiv U^{-1}(M^T D^{-1} M)$. The second right-eigenvector of $\mathcal{P}$ then defines the (unstandardized) PCIs. Similarly, collapsing the product dimension leads to the $n_c \times n_c$ row-normalized stochastic matrix $\mathcal{C} \equiv D^{-1}(MU^{-1}M^T)$. The second right-eigenvector of $\mathcal{C}$ then defines the (unstandardized) ECIs. Often, the PCIs and ECIs are standardized to have zero mean and unit variance across products or countries[27]. However, we discuss the unstandardized vectors, and indeed our theory implies that standardizing removes information, a point revisited in the Supplementary Information.

Relationship to the PCI and ECI. To see that elements of $\mathbf{v}_2$ are related to PCIs, note that the structure of any eigenvector of our network of activities depends on how we construct this network, and in particular on how we define proximity between activities. A variety of measures of proximity have been used, and this creates ambiguity about the nature of the eigenvectors. However, it turns out that the structure of the second eigenvector is not sensitive to the particular way proximity is measured.

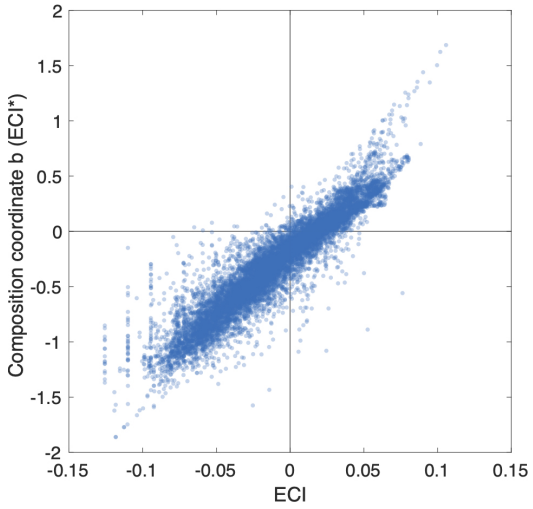


FIG. 3. The b coordinate (ECI*) versus ECI for 249 countries and regions and 57 years computed from Comtrade data[36].

In fact, for one specific choice of the proximity matrix, the vector $\mathbf{v}_2$ and the vector of PCIs (**PCI**) are identical. This occurs when the proximities are taken to be

$$\Phi^P_{pp'} \equiv \sum_c \frac{M_{cp} M_{cp'}}{D_c}. \tag{10}$$

In this case, it is straightforward to show (Supplementary Information) that $\mathbf{v}_2$ and **PCI** are second eigenvectors satisfying the same eigenvector equation, and are therefore identical up to an arbitrary proportionality constant.

However, the resemblance of $\mathbf{v}_2$ with **PCI** is not a quirk of this special choice of proximity matrix. In general, other proximity measures resemble Φ^P and their second eigenvectors strongly resemble one another and **PCI** (Supplementary Information). For example, suppose the network of our dynamical model is described using the Φ^M matrix noted in the Background (a measure of proximity used to define the Product Space[15]). The resemblance between Φ^M and Φ^P is strong, and the elements of the eigenvector $\mathbf{v}_2$ of the resulting Laplacian matrix L_Φ are highly correlated to those of **PCI**, with a Pearson correlation 0.947 and a Spearman rank correlation 0.950. To draw a distinction between the conventional PCI vector and the second eigenvector of the dynamical model, we call the latter PCI*: $\mathbf{PCI}^* \equiv \mathbf{v}_2$. (I.e. we tie the definition of $\mathbf{PCI}^*$ to the proximity matrix Φ used by the dynamical model.) In general these vectors are strongly correlated, and in the special case $\Phi = \Phi^P$ they are identical, $\mathbf{PCI}^* = \mathbf{PCI}$.

The coordinate $b(t)$ pinpoints where a country lies on the axis of economic change described by $\mathbf{PCI}^*$. The dynamical model invites us to see this coordinate as capturing a direction of activity changes. However, this coordinate is also closely related to the complexity metric known as the Economic Complexity Index. This suggests that we can define an ECI-like quantity in our framework: $\mathrm{ECI}^*(t) \equiv b(t)$. To see that this coordinate is in fact closely related to the ECI, note that a country's ECI can be computed[45,46] as the average PCI of the products in which the country has an RCA greater than 1:

$$\mathrm{ECI}_c = \sum_p \left(\frac{M_{cp}}{\sum_{p'} M_{cp'}} \right) \mathrm{PCI}_p. \tag{11}$$

The ECI* can be computed by observing that $b(t) = c_2(t)/c_1(t)$, and then using Eq. (6) to obtain the eigenexpansion coefficients c_2 and c_1. The resulting calculation is done in the Supplementary Information and leads to

$$\mathrm{ECI}^*_c = \sum_p \left(\frac{R_{cp}\pi_p}{\sum_{p'} R_{cp'}\pi_{p'}} \right) \mathrm{PCI}^*_p. \tag{12}$$

Thus, the ECI and $b(t)$ are simply different averages of PCIs or PCI*s. The conventional ECI gives equal weight to activities in which the country has an RCA greater than 1, and zero weight to other activities. The ECI* weights activities by their RCA, adjusted by the elements of the first left eigenvector $\boldsymbol{\pi} = \mathbf{w}_1$. Not surprisingly, these averages are strongly correlated for any proximity matrices we consider (see Supplementary Information).

Notably, while the $\mathbf{PCI}^*$ and **PCI** coincide in one special case ($\Phi = \Phi^P$), the same is never true of the ECI* and ECI. Differences between these quantities will always remain because of the different weighting schemes. In the closest approach (i.e. $\Phi = \Phi^P$), the Pearson correlation between the ECI* and the ECI is 0.919, and the Spearman rank correlation is 0.902 (Fig. 3). As a comparison, when $\Phi = \Phi^M$, we observe a Pearson correlation of 0.835 and Spearman rank correlation of 0.846. In the Supplementary Information, we discuss why the weights used to calculate ECI* may have certain desirable properties.

We thus arrived at quantities, PCI* and ECI*, that closely resemble the PCI and ECI. Yet the mathematical route we took leads to important differences in how we should interpret the PCI* and ECI*. The conventional PCI and ECI are defined in mathematically symmetric terms, and are interpreted symmetrically (e.g. as complexity measures[20], or as abstract scores that cluster products or countries by their similarity[45]). But we would view the doppelgängers of these quantities – the $\mathbf{PCI}^*$ and ECI* – in very different terms. First, the $\mathbf{PCI}^*$ and ECI* differ fundamentally from one another: The PCI*s, collectively, define a pattern or axis of economic change in the vector space of activities, while the ECI*s individually convey where different countries lie on this axis.

Second, the $\mathbf{PCI}^*$ and ECI* are not simultaneously determined, i.e. by a system of coupled equations[20,21,26]. Instead, they arise from math with a clear sequence (Fig. 4a-c): There is a diversification process that is shaped by the relatedness network given by Φ, from which one eigenvector corresponds to $\mathbf{PCI}^*$, onto which a given region's coordinate projection ECI* can be computed.

Third, correspondingly, the relevance of $\mathbf{PCI}^*$ and ECI* is not tied to considerations of how to infer complexity from data, but to how aptly the vector of PCI*s

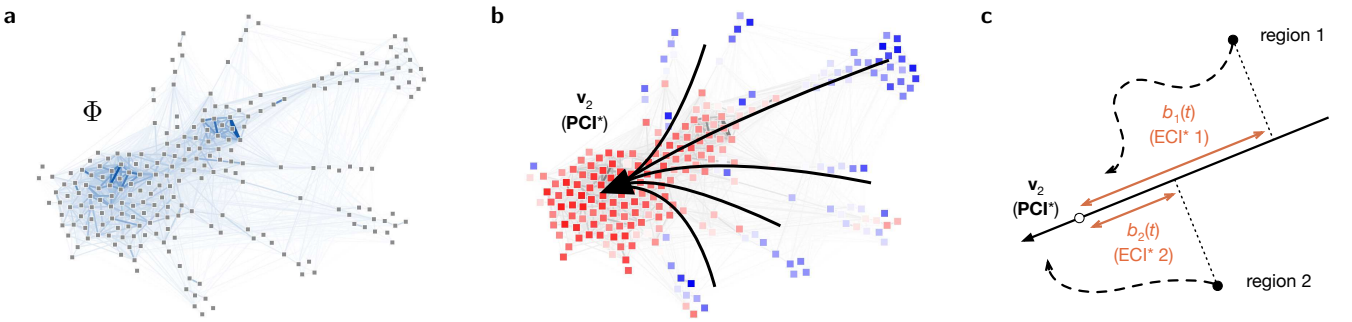


FIG. 4. The order in which key quantities are computed in our framework. **a** A relatedness network is first defined, on which various patterns of transitions can be described, such as **b** the compositional mode of change described by the elements of $\mathbf{v}_2$ (**PCI**$*$). Negative values (blue) correspond to economic activities that are vacated in relative terms, while positive values (red) correspond to activities are reached. **c** Finally, a region's coordinate projection onto this mode determines its value of b (ECI*).

describes a long-lived direction of change on the network. Thus, even though the closest approach to the conventional metrics occurs when Φ is taken to be Φ^P, there is no inherent requirement to assume this. Proximities between activities can be measured with a range of existing or new measures, and could be selected based on the quality of forecasts of future transitions in activities. Improving their predictions then improves the estimates of the **PCI*** and ECI*, offering a path for refinement.

Dimensionality reduction. Recent works have approached economic complexity analysis through the framework of dimensionality reduction. Mealy et al. (2019)[45] point out that the conventional ECI can be seen as a dimensionality-reduction tool, being equivalent to a spectral clustering algorithm that orders regions along an axis that scores the similarity in their baskets of activities. Sciarra et al. (2020)[26] put forward a complexity index that reduces the dimensionality of economic activity data in a different way, combining information from two eigenvectors of a country similarity graph. Brummitt et al. (2020)[19] apply machine learning methods to historical export data to extract dimensions that characterize variation in export baskets across countries and time, finding axes that strongly correlate with the PCIs. While we have emphasized the dynamical meaning of $\mathbf{v}_2$ (**PCI***) and $b(t)$ (ECI*), one could also think about these quantities in a dimensionality-reduction framework.

First, the outcome of eigenmode decomposition resembles dimensionality-reduction techniques in that it returns axes (the eigenvectors $\mathbf{v}_\mu$) in which one can describe high-dimensional data (the matrix of activity baskets $\mathbf{R}(t)$), along with the coordinate projections of each data point onto these axes (the coefficients $c_\mu(t)$). Differences arise in how directions in the vector space are determined and interpreted, as dimensionality-reduction techniques often aim to identify directions that capture high amounts of variation in data, while eigenmodes characterize coherent patterns in dynamics.

In fact, an eigenmode decomposition not only generates a *representation* of data in a low-dimensional vector space, but shows that the model predicts that countries' activity baskets will actually *converge* to such a subspace (i.e. corresponding to the directions with the largest time scales). Differences between countries shrink first among quickly-decaying dimensions, leaving activity baskets scattered primarily along slowly-decaying ones, a well-known outcome of the interaction of structure and dynamics in networks[47,48]. (See also Supplementary Information.)

The ECI, country Fitness, Production Ability, and GENEPY. Finally, we expand our scope, and consider how the structural change coordinates A and b compare with complexity metrics more generally. We find that existing complexity metrics each have an approximate correspondence to one or both of these two coordinates. We have already made clear such a correspondence between the ECI and the b coordinate. Another complexity metric, country Fitness[21], is highly correlated with the A coordinate (Pearson correlation 0.74, Spearman rank correlation 0.87). A third metric, Production Ability[25], is also highly correlated with the A coordinate (Pearson correlation 0.90, Spearman correlation 0.91). Finally, a metric called GENEPY, recently proposed by Sciarra et al.[26], is connected to both coordinates. GENEPY is a hybrid metric that combines information from two component measures $X_{c,1}$ and $X_{c,2}$. By design, $X_{c,1}$ is closely related to Fitness and $X_{c,2}$ is closely related to the ECI, so that under certain conditions $X_{c,1}d_c$ corresponds to Fitness and $X_{c,2}/\sqrt{d_c}$ corresponds to the ECI, where d_c is country c's diversity[26]. Similarly, $X_{c,1}d_c$ and $X_{c,2}/\sqrt{d_c}$ are empirically strongly correlated with the A and b coordinates respectively (Pearson correlations 0.77 and 0.77, Spearman correlations 0.89 and 0.81, respectively). (See also Supplementary Information for further discussion.)

Fig. 5 shows that complexity metrics and related measures fall into two groups that correlate strongly with the two coordinates expected by our theory. One is a group of diversity-like quantities, which include diversity d_c itself, the A coordinate of our theory, country Fitness, Production Ability, and the quantity $X_{c,1}d_c$. The other is a group of quantities that reflect compositional aspects of an economy's activity mix, which include the ECI, the b coordinate of our theory (ECI*), and $X_{2,c}/\sqrt{d_c}$. The

GENEPY metric, which combines information associated with both groups, does not fall clearly in either one. Through the analytical lens of this paper, complexity metrics emphasize different aspects of the same development process, related either to changes in the diversity of economic activities, or to changes in their composition. Yet as Fig. 2 illustrated, and both classic and recent literature supports[18,19,42,43], in practice both types of changes are important characteristics of the process of development.

III. DISCUSSION

We offer two broad ways to view the results here. In general, our theory represents a mathematical framework that bridges short-term and long-term descriptions of structural change. The Principle of Relatedness and its empirical implementations describe structural change on short, year-to-year time scales, and across fine-grained sectors of activities. In contrast, classic observations of structural change[42,43] emphasize the slow transition of economies over decades across broad economic sectors, as well as changes in an economy's diversity[18]. As one would hope, a dynamical framework based on short-term, sectorally-disaggregated observations is consistent with and bridges into classic observations of long-term change.

We can also relate the results here to recent work in the field of economic complexity. Our theory connects two hitherto unconnected branches of this field – studies of relatedness, and complexity metrics – in one framework. Essentially, these branches refer to short- and long-term consequences of the same assumptions about economic development. The long-term compositional mode of change in our theory is closely related to the complexity metrics PCI and ECI. This is surprising, as these metrics have been motivated for reasons conceptually removed from those we leverage here. Had these metrics not been proposed as ways to infer complexity, very similar metrics could have been motivated from empirical work on the Principle of Relatedness using a standard eigenmode decomposition, as coordinates describing a particular pattern of economic transformation.[2] Our results also highlight that different complexity metrics in general emphasize different aspects of the same development process. In this way, our emphasis contrasts with the bulk of discussions of complexity metrics, which have focused on the ability of these metrics to infer 'complexity'[20–26]. Our results do not directly support nor contradict the interpretation of complexity metrics as inference methods. On the other hand, they do give these indices novel motivation and support their empirical relevance. Furthermore, we can also exploit the novel conceptual foundations of our approach to suggest new ways of using such metrics to quantify structural change. These extensions are discussed in the Supplementary Information.

Many questions remain about the factors that drive or inhibit the development of economies. We explored our results in the setting of country exports, but similar analyses could be carried out in networks of occupations[49], industry output[50,51], or even in networks of related locations[16], where the roles of locations and activities would be reversed by projecting the bipartite network of locations and activities onto location nodes instead of activities. We suggest a framework like the one here can be a step toward improved theoretical models that are closely tied with empirical metrics, shedding light on the determinants of growth and development.

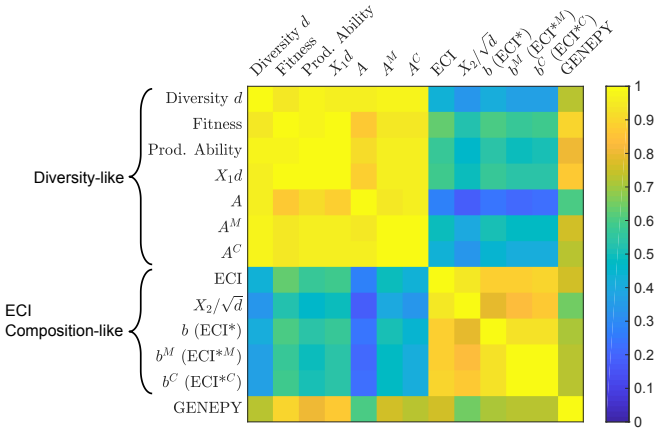


FIG. 5. Spearman rank correlations among our A and b (ECI*) coordinates, several proposed measures of complexity (ECI, Fitness, Production Ability, and GENEPY), and related quantities. The coordinate A is the average ability coordinate used throughout the text defined using the proximity matrix Φ^P, while A^M and A^C refer to the same coordinate obtained from using two alternative proximity matrices Φ^M (Eq. (S43)) and Φ^C (Eq. (S44)). Similarly ECI* is the b coordinate used throughout the text and ECI*^{M} and ECI*^{C} are the same coordinate using Φ^M or Φ^C. X_1 and X_2 are the first and second components of the GENEPY metric. All quantities were computed for year 2016 using UN Comtrade data[36].

MATERIALS AND METHODS

Description of data We use cleaned UN Comtrade data[36] publicly available at Harvard Dataverse[52]. Trade data are reported twice: once as exports by the exporting country and once as imports by the importing country. The data cleaning corrects trade flows to increase the consistency between importer and exporter records of the same flow, and further corrects reported values using an index of reliability based on the consistency of reported values over time. The data provides total export volumes of products for 249 regions over the period 1962 - 2018 (11,544 export baskets observed across all regions and

[2] Relatedly, a feature of the ECI that has been highlighted since its introduction is its ability to forecast economic growth. A possibility raised by our theory is that the ECI acquires some of this forecasting capacity through its ability to describe a common long-term pattern of activity change induced by the network structure of relatedness.

years). We analyze these data at the 3-digit SITC product level (235 exported products).

Calculation of relatedness network To obtain our results in Fig. 2 we first compute the matrix of relatedness between products as $\Phi^P = M^T D^{-1} M$ (Eq. (10)). We fix Φ in the initial year of our data (1962) in accordance with the idea, inherent in the PoR, that economies diversify across a fixed (or at least slow-moving) space of related activities. From Φ we compute the row-normalized version of this matrix $\tilde{\Phi}_{pp'} = \Phi_{pp'} / \sum_{p''} \Phi_{pp''}$ as described in the text, whose right ($\mathbf{v}_\mu$) and left ($\mathbf{w}_\mu$) eigenvectors are the basis for the remainder of the analysis. We normalize these eigenvectors with a weighted L_2-norm described in Supplementary Information, though this choice of norm is just a convention and is not consequential, as different choices will simply rescale elements of $\mathbf{v}_\mu$ or $\mathbf{w}_\mu$ and thus rescale the coordinates projected onto the eigenvectors.

Calculation of R, A coordinate, and b coordinate For each country c, product p, and year we computed the Balassa index of revealed comparative advantage (RCA) $R_{cp} = (X_{cp} / \sum_{p'} X_{cp'}) / (\sum_{c'} X_{c'p} / \sum_{c',p'} X_{c'p'})$ where X_{cp} is the value of c's exports in product p. RCAs are heavy tailed and are commonly transformed to weaken the effect of extreme values. Here, we transform them using the function $g(R) = \alpha \log(1 + R/R_0)$. The constant $\alpha = [\log(1 + 1/R_0)]^{-1}$ is included for convenience so that an RCA of 1 is mapped to 1 on the transformed scale. This function behaves linearly for small values of R and logarithmically for large values, and thus achieves the goal of weakening the effect of very large RCAs while also handling RCAs that are identically zero (i.e. the many instances where a country has no exports of a product). The parameter R_0 (which sets the transition between the linear and logarithmic regimes) is tuned to maximize the amount of variance in export baskets that the b coordinate explains across time and countries. See the Supplementary Information for a detailed discussion. Our results can be reproduced using our obtained value $R_0 = 0.115$. We then calculate for each country c and time t the RCA vector $\mathbf{R}_c(t)$ on this transformed scale and use these vectors to compute the A and b coordinates. The A coordinate for country c and time t is computed as $A_c(t) = \boldsymbol{\pi}^T \mathbf{R}_c(t)$, where $\boldsymbol{\pi} = \mathbf{w}_1$ is the first left eigenvector. The b coordinate is computed as $b_c(t) = \mathbf{w}_2^T \mathbf{R}_c(t) / \mathbf{w}_1^T \mathbf{R}_c(t)$, which is equivalent to Eq. (12), as noted in the main text.

ACKNOWLEDGEMENTS

We thank Ulrich Schetter, R. Maria del Rio Chanona, and Ricardo Hausmann for valuable discussions. Frank Neffke acknowledges financial support from the Austrian Research Agency (FFG), project #873927 (ESSENCSE).

[1] Hausmann, R., Hwang, J. & Rodrik, D. What you export matters. *J. Econ. Growth* **12**, 1–25 (2007).
[2] Porter, M. The economic performance of regions. *Regional Studies* **37**, 549–578 (2003).
[3] Delgado, M., Porter, M. E. & Stern, S. Clusters, convergence, and economic performance. *Research Policy* **43**, 1785–1799 (2014).
[4] Gathmann, C. & Schönberg, U. How general is human capital? a task-based approach. *Journal of Labor Economics* **28**, 1–49 (2010).
[5] Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M. & Hidalgo, C. A. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* **109**, 1695–1709 (2016).
[6] Farjoun, M. Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. *Organization Science* **5**, 185–199 (1994).
[7] Lien, L. B. & Klein, P. G. Using competition to measure relatedness. *Journal of Management* **35**, 1078–1107 (2009).
[8] Neffke, F. & Henning, M. Skill relatedness and firm diversification. *Strategic Management Journal* **34**, 297–316 (2013).
[9] Neffke, F., Henning, M. & Boschma, R. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography* **87** (2011).
[10] Boschma, R., Minondo, A. & Navarro, M. Related variety and regional growth in Spain. *Papers in Regional Science* **91** (2012).
[11] Boschma, R., Minondo, A. & Navarro, M. The emergence of new industries at the regional level in Spain: A proximity approach based on product relatedness. *Economic geography* **89**, 29–51 (2013).
[12] Essletzbichler, J. Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Regional Studies* **49**, 752–766 (2015).
[13] Zhu, S., He, C. & Zhou, Y. How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography* **17**, 521–545 (2017).
[14] Hausmann, R. & Klinger, B. The structure of the product space and the evolution of comparative advantage. *CID Working Paper Series* (2007).
[15] Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317** (2007).
[16] Hausmann, R., Stock, D. P. & Yıldırım, M. A. Implied comparative advantage. *Research Policy* 104143 (2021).
[17] Hidalgo, C. A. *et al.* The principle of relatedness. In *International conference on complex systems*, 451–457 (Springer, 2018).
[18] Imbs, J. & Wacziarg, R. Stages of diversification. *American Economic Review* **93**, 63–86 (2003).
[19] Brummitt, C. D., Gómez-Liévano, A., Hausmann, R. & Bonds, M. H. Machine-learned patterns suggest that diversification drives economic development. *Journal of the Royal Society Interface* **17**, 20190283 (2020).
[20] Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proceedings of the National*

Academy of Sciences **106**, 10570 – 10575 (2009).

[21] Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Scientific Reports* **2**, 1–7 (2012).

[22] Kemp-Benedict, E. An interpretation and critique of the method of reflections. *Munich Personal RePEc Archive (MPRA)* (2014).

[23] Mariani, M. S., Vidmer, A., Medo, M. & Zhang, Y.-C. Measuring economic complexity of countries and products: which metric to use? *The European Physical Journal B* **88**, 1–9 (2015).

[24] Morrison, G. *et al.* On economic complexity and the fitness of nations. *Scientific Reports* **7**, 1–11 (2017).

[25] Bustos, S. & Yıldırım, M. A. Production ability and economic growth. *Research Policy* 104153 (2020).

[26] Sciarra, C., Chiarotti, G., Ridolfi, L. & Laio, F. Reconciling contrasting views on economic complexity. *Nature Communications* **11**, 1–10 (2020).

[27] Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M. & Simoes, A. *The atlas of economic complexity: Mapping paths to prosperity* (MIT Press, 2014).

[28] Caldarelli, G. *et al.* A network analysis of countries' export flows: Firm grounds for the building blocks of the economy. *PloS one* **7**, e47278 (2012).

[29] Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G. & Pietronero, L. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one* **8**, e70726 (2013).

[30] Albeaik, S., Kaltenberg, M., Alsaleh, M. & Hidalgo, C. A. Improving the economic complexity index. *arXiv preprint arXiv:1707.05826* (2017).

[31] Balassa, B. Trade liberalisation and "revealed" comparative advantage 1. *The Manchester School* **33**, 99–123 (1965).

[32] Hillman, A. L. Observations on the relation between "revealed comparative advantage" and comparative advantage as indicated by pre-trade relative prices. *Review of World Economics* **116**, 315–321 (1980).

[33] Hinloopen, J. & van Marrewijk, C. Empirical relevance of the hillman condition for revealed comparative advantage: 10 stylized facts. *Applied Economics* **40**, 2313–2328 (2008).

[34] Boschma, R., Balland, P.-A. & Kogler, D. F. Relatedness and technological change in cities: The rise and fall of technological knowledge in us metropolitan areas from 1981 to 2010. *Industrial and Corporate Change* **24**, 223–250 (2015).

[35] Petralia, S., Balland, P.-A. & Morrison, A. Climbing the ladder of technological development. *Research Policy* **46**, 956–969 (2017).

[36] United Nations. UN Comtrade International Trade Statistics Database. URL `https://comtrade.un.org/`.

[37] Hoen, A. R. & Oosterhaven, J. On the measurement of comparative advantage. *The Annals of Regional Science* **40**, 677–691 (2006).

[38] Elekes, Z., Boschma, R. & Lengyel, B. Foreign-owned firms as agents of structural change in regions. *Regional Studies* **53**, 1603–1613 (2019).

[39] United Nations. Un list of least developed countries. URL `https://unctad.org/topic/least-developed-countries/list`.

[40] Hines Jr, J. R. Treasure islands. *Journal of Economic Perspectives* **24**, 103–26 (2010).

[41] The World Bank. World Bank Development Indicators (2018). URL `http://data.worldbank.org/`.

[42] Clark, C. *The conditions of economic progress.* (London: Oxford, 1967).

[43] Kuznets, S. Quantitative aspects of the economic growth of nations: II. Industrial distribution of national product and labor force. *Econ. Dev. Cult. Change* **5**, 1–111 (1957).

[44] Cadot, O., Carrère, C. & Strauss-Kahn, V. Export diversification: What's behind the hump? *Review of Economics and Statistics* **93**, 590–605 (2011).

[45] Mealy, P., Farmer, J. D. & Teytelboym, A. Interpreting economic complexity. *Science Advances* **5**, eaau1705 (2019).

[46] Hill, M. O. Reciprocal averaging: An eigenvector method of ordination. *The Journal of Ecology* 237–249 (1973).

[47] Simon, H. A. & Ando, A. Aggregation of variables in dynamic systems. *Econometrica: Journal of the Econometric Society* 111–138 (1961).

[48] Schaub, M. T., Delvenne, J.-C., Lambiotte, R. & Barahona, M. Structured networks and coarse-grained descriptions: A dynamical perspective. *Advances in Network Clustering and Blockmodeling* 333–361 (2019).

[49] Neffke, F. M., Otto, A. & Weyh, A. Inter-industry labor flows. *Journal of Economic Behavior & Organization* **142**, 275–292 (2017).

[50] Neffke, F. M. The value of complementary co-workers. *Science Advances* **5**, eaax3370 (2019).

[51] Hong, I., Frank, M. R., Rahwan, I., Jung, W.-S. & Youn, H. The universal pathway to innovative urban economies. *Science Advances* **6**, eaba4934 (2020).

[52] The Growth Lab at Harvard University. International trade data (SITC, Rev. 2) v4.0 (2019). URL `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/H8SFD2&version=4.0`.

arXiv:2110.09673v1 [physics.soc-ph] 19 Oct 2021

Bridging the short-term and long-term dynamics of economic structural change (Supplementary Information)

James McNerney, Yang Li, Andrés Gómez-Liévano, Frank Neffke

Contents

S1 Supplementary theoretical results and discussion **2**
S1.1 Principle of Relatedness as a dynamical model . . . 2
S1.2 Eigenmode decomposition of model . . . 3
S1.3 Additional discussion of the eigenvectors of L_Φ . . . 4
S1.3.1 Cancellation of imaginary components in complex conjugate pairs . . . 4
S1.3.2 Relationship between left and right eigenvectors . . . 4
S1.3.3 Normalization of eigenvectors . . . 5
S1.4 The A coordinate . . . 6
S1.5 Review of the PCI and ECI . . . 7
S1.6 The PCI* and ECI* . . . 8
S1.7 Dimensionality reduction . . . 11
S1.8 Possible extensions to structural change coordinates . . . 13

S2 Supplementary empirical results **14**
S2.1 Transformation of RCAs . . . 14
S2.2 Comparison of PCI*/PCI and ECI*/ECI across proximity measures . . . 16
S2.3 Individual country trajectories . . . 18
S2.4 Correlations of coordinates with complexity metrics . . . 19

S1 Supplementary theoretical results and discussion

S1.1 Principle of Relatedness as a dynamical model

For convenience, this section and the next review the model of the main text and its eigenmode decomposition. Following models in the literature, we suppose that diversification to new activities can be described as a process of spreading on a network of activities. A region's levels of ability in different activities are given by the elements of a vector $\mathbf{R}(t)$, which evolve as

$$\dot{\mathbf{R}}(t) = \gamma(t)\mathbf{R}(t) - \frac{1}{\tau}L_\Phi \mathbf{R}(t), \tag{S1}$$

where $\gamma(t)$ is an arbitrary growth rate, L_Φ is a graph Laplacian, and τ is a time scale. The terms in Eq. (S1) capture two kinds of changes: A region's abilities can generally rise or fall as a whole, and they can shift to weight activities differently, changing which ones receive the most emphasis.

The Laplacian L_Φ captures the structure of the network on which this spreading process occurs. We let $L_\Phi = I - \tilde{\Phi}$, where $\tilde{\Phi}$ is a proximity matrix with rows normalized to sum to 1, i.e. $\tilde{\Phi}_{pp'} = \Phi_{pp'} / \sum_{p''} \Phi_{pp''}$. With this assumption, in index form, a discrete-time approximation of Eq. (S1) reads

$$\frac{R_p(t+\Delta t) - R_p(t)}{\Delta t} \approx \gamma(t) R_p(t) + \frac{1}{\tau}\left(\sum_{p'} \tilde{\Phi}_{pp'} R_{p'}(t) - R_p(t)\right). \tag{S2}$$

In this form our dynamical model can be more easily compared with regression setups that implement the PoR in the literature. Ability in activity p changes between periods t and $t + \Delta t$ in a way that depends on ability in nearby products. When the average ability in activities around p ($\sum_{p'} \tilde{\Phi}_{pp'} R_{p'}(t)$) exceeds that in p itself ($R_p(t)$), ability in p rises. Abilities can also rise or fall as a whole, and this is captured by the term $\gamma(t)R_p(t)$.

Integrating Eq. (S1) over time leads to

$$\mathbf{R}(t) = A(t)P(t)\mathbf{R}(0), \tag{S3}$$

where the dynamics captured by the two terms in Eq. (S1) are now captured by the scalar pre-factor $A(t)$ and the matrix $P(t)$. The prefactor $A(t) \equiv e^{\int_0^t \ (s)\, ds}$ is a shift factor that captures the accumulated growth in overall ability up to time t, scaling the elements of $\mathbf{R}$ up or down as a whole. The matrix $P(t) = e^{-L_\Phi(t/\tau)}$ is a stochastic matrix that transforms $\mathbf{R}$, changing the relative emphasis on different activities over time.

S1.2 Eigenmode decomposition of model

Let $\mathbf{v}_\mu$ be the μth right eigenvector of L_Φ and let κ_μ be its eigenvalue. Ordering eigenvalues from least to greatest, the first right eigenvector has $\kappa_1 = 0$ and is a uniform vector of real positive numbers. It can be taken to be a vectors of 1s, $\mathbf{v}_1 = \mathbf{1}$. Higher-order eigenvectors ($\mathbf{v}_2$ and up) have $\kappa_\mu > 0$, and contain a mix of elements with positive and negative real parts. Some eigenvectors may contain imaginary parts, but this poses no special difficulties; see Section S1.3.1.

The right eigenvectors of L_Φ form a basis for the vector space of ability vectors $\mathbf{R}(t)$, and as a result, one can write any such vector as a linear combination of the eigenvectors:

$$\mathbf{R}(t) = \sum_\mu c_\mu(t)\mathbf{v}_\mu. \tag{S4}$$

Letting $\mathbf{w}_\mu$ be the μth *left* eigenvector L_Φ, the coefficients $c_\mu(t)$ above may be computed by exploiting the biorthogonality of L_Φ's left and right eigenvectors, $\mathbf{w}_\mu^\dagger \mathbf{v}_\nu = \delta_{\mu\nu}$:

$$c_\mu(t) = \mathbf{w}_\mu^\dagger \mathbf{R}(t). \tag{S5}$$

Proceeding with the usual steps of eigendecomposition, we decompose $\mathbf{R}(t)$ to separate different modes of change in $\mathbf{R}(t)$ according to their time scales. Plugging Eq. (S4) into Eq. (S3) gives $\mathbf{R}(t) = A(t)\sum_\mu c_\mu(0)e^{-\kappa_\mu t/\tau}\mathbf{v}_\mu$. The first term contains the first eigenvector $\mathbf{v}_1 = \mathbf{1}$ and corresponds to the fixed point of the simpler model $\mathbf{R}(t) = P(t)\mathbf{R}(0)$ (i.e. the model with the factor $A(t)$ fixed at 1). Separating the first term from other terms of the sum we have

$$\mathbf{R}(t) = A(t)\left(c_1(0)\mathbf{1} + \sum_{\mu\geq 2} c_\mu(0)e^{-t/\tau_\mu}\mathbf{v}_\mu\right), \tag{S6}$$

where we defined $\tau_\mu \equiv \tau/\kappa_\mu$.

Eq. (S6) describes the following behavior. First, neglecting the effect of the shift factor $A(t)$, the activity basket of a region eventually converges to a uniform vector $c_1(0)\mathbf{1}$. In this state, the region has equal ability in all products. On its way to this state, the basket shows higher ability in some products and lower ability in others. Each eigenvector $\mathbf{v}_\mu$ describes a different pattern of deviations from uniformity, each decaying with time at the rate set by τ_μ, the characteristic time scale of the μth eigenmode. If the pre-factor $A(t)$ were fixed and equal to 1, then a region would be destined to have ability $c_1(0)$ in every product. However, letting $A(t)$ change over time, the final level of ability in products can be arbitrarily high or low.

We now focus on the first two eigenvectors of the expansion. Absorbing the coefficient $c_1(0)$ into $A(t)$, and defining the coefficient $b(t) \equiv c_2(0)e^{-t/\tau_2}$, the longest-lived dynamics of $\mathbf{R}(t)$ are described by the two leading terms of the eigenexpansion:

$$\mathbf{R}(t) = A(t)\mathbf{1} + A(t)b(t)\mathbf{v}_2 + \cdots. \tag{S7}$$

The coefficients $A(t)$ and $b(t)$ each have different effects. A shift in $A(t)$ corresponds to a region realizing a level change in its abilities across activities. A shift in $b(t)$ corresponds to a compositional shift. Some activities rise in ability and others fall, as determined by the signs and magnitudes of entries in $\mathbf{v}_2$. Together, these coordinates situate the activity basket $\mathbf{R}(t)$ in a 2D space.

The two terms of Eq. (S7) are further discussed in Sections S1.4 and S1.6. Section S1.4 shows how the $A(t)$ coordinate is related to an average of the elements of $\mathbf{R}(t)$. Section S1.6 shows how the $b(t)$ coordinate is related to the ECI.

S1.3 Additional discussion of the eigenvectors of L_Φ

Here we provide additional discussion of the left and right eigenvectors of the Laplacian matrix L_Φ. The first section reviews how complex eigenvectors could be handled by summing them in pairs. The second section derives an important relationship between the left and right eigenvectors. We later exploit this relationship to show that the ECI* closely resembles the conventional ECI in functional form. The third section describes our convention for normalizing the left and right eigenvectors.

S1.3.1 Cancellation of imaginary components in complex conjugate pairs

The eigenvectors and eigenvalues of the graph Laplacian L_Φ in general are complex, but this poses no special difficulties in interpreting the eigenmodes described by the vectors $\mathbf{v}_\mu$ because all eigenvectors with imaginary elements occur in complex conjugate pairs. In practice, neither of the two eigenvectors that we focus on here contained imaginary components, but for completeness, we discuss how such cases could be handled. As noted in Eq. (S5), the projection of region c onto the μth eigenmode is given by $c_\mu(t) = \mathbf{w}_\mu^\dagger \mathbf{R}(t)$. If the elements of $\mathbf{w}_\mu = \mathbf{a}_\mu + i\mathbf{b}_\mu$ contain imaginary parts, then it has a paired eigenvector $\mathbf{w}_{\mu'} = \mathbf{w}_\mu^* = \mathbf{a}_\mu - i\mathbf{b}_\mu$. Wherever this is the case we may combine the eigenvectors and consider the coefficients in pairs, i.e.

$$c_\mu(t) + c_{\mu'}(t) = \mathbf{w}_\mu^\dagger \mathbf{R}(t) + \mathbf{w}_{\mu'}^\dagger \mathbf{R}(t) = \left[\mathbf{w}_\mu^\dagger + (\mathbf{w}_\mu^\dagger)^*\right] \mathbf{R}(t) \tag{S8}$$

for which the summed vector $\mathbf{w}_\mu^\dagger + (\mathbf{w}_\mu^\dagger)^*$ (and therefore the combined projection onto both modes, $c_\mu(t) + c_{\mu'}(t)$) is manifestly real.

S1.3.2 Relationship between left and right eigenvectors

Here we derive a relationship between the left and right eigenvectors of L_Φ, Eq. (S9). We will use this relationship later to derive an expression for the ECI*, Eq. (S32).

First, note that $L_\Phi = I - \tilde{\Phi}$ and $\tilde{\Phi}$ share the same eigenvectors. The matrix $\tilde{\Phi} = D^{-1}\Phi$ with $D = \mathsf{Diag}(\Phi\mathbf{1})$ is a row-normalized stochastic matrix, and so its principal right eigenvector may be taken to be a vector of 1s, $\mathbf{v}_1 = \mathbf{1}$. Its principal left eigenvector can be understood as a vector of stationary probabilities $\boldsymbol{\pi}$ for the Markov chain described by $\tilde{\Phi}$, $\mathbf{w}_1 = \boldsymbol{\pi}$.

Define $\Pi = \mathsf{Diag}(\boldsymbol{\pi})$. If Φ is symmetric, then the left and right eigenvectors of $\tilde{\Phi}$ are related by

$$\Pi \mathbf{v}_\mu = \alpha \mathbf{w}_\mu \tag{S9}$$

where α is a constant that depends on the normalization of the eigenvectors.

Proof: By definition the right eigenvectors of $\tilde{\Phi}$ satisfy $\tilde{\Phi}\mathbf{v}_\mu = \lambda_\mu \mathbf{v}_\mu$, which can be rearranged as

$$(D^{-1/2}\Phi D^{-1/2})(D^{1/2}\mathbf{v}_\mu) = \lambda_\mu (D^{1/2}\mathbf{v}_\mu).$$

The vector $D^{1/2}\mathbf{v}_\mu$ is thus a right eigenvector of the symmetric matrix $Q \equiv D^{-1/2}\Phi D^{-1/2} = D^{1/2}\tilde{\Phi}D^{-1/2}$. Similarly, the left eigenvectors of $\tilde{\Phi}$ satisfy $\mathbf{w}_\mu^\dagger \tilde{\Phi} = \lambda_\mu \mathbf{w}_\mu^\dagger$, which can rearranged as

$$(\mathbf{w}_\mu^\dagger D^{-1/2})(D^{-1/2}\Phi D^{-1/2}) = \lambda_\mu (\mathbf{w}_\mu^\dagger D^{-1/2}),$$

showing that $D^{-1/2}\mathbf{w}_\mu$ is a left eigenvector of the same matrix Q. Finally, since Q is symmetric, any right eigenvector is also a left eigenvector, and thus

$$D^{1/2}\mathbf{v}_\mu = \alpha D^{-1/2}\mathbf{w}_\mu$$

for some constant α. We therefore have $D\mathbf{v}_\mu = \alpha \mathbf{w}_\mu$. In particular, the first left and right eigenvectors of $\tilde{\Phi}$ are $\mathbf{v}_1 \propto \mathbf{1}$ and $\mathbf{w}_1 \propto \boldsymbol{\pi}$, and so we have $D\mathbf{1} \propto \boldsymbol{\pi}$. Since D is diagonal, in index form this reads $D_{pp} \propto \pi_p$, i.e. D is a matrix that is proportional to Π, giving us Eq. (S9).

S1.3.3 Normalization of eigenvectors

Here we describe our convention for normalizing left and right eigenvectors. This convention helps set the numerical scale of the ECI* (and in principle other coordinate projections).

Left and right eigenvectors are in general biorthogonal, obeying $W^T V = \Psi$ for some diagonal matrix Ψ, and it simplifies many calculations to require that $\Psi = I$, so that left and right eigenvectors obey

$$\mathbf{w}_\nu^\dagger \mathbf{v}_\mu = \delta_{\nu\mu}. \tag{S10}$$

This condition does not determine a normalization for the $\mathbf{w}_\mu$ and $\mathbf{v}_\mu$ eigenvectors, since these can be rescaled as $\mathbf{v}_\mu \to c\mathbf{v}_\mu$ and $\mathbf{w}_\mu \to \frac{1}{c}\mathbf{w}_\mu$ while still preserving Eq. (S10). However, we can pin down a normalization for the eigenvectors by requiring that they also satisfy Eq. (S9) with proportionality constant $\alpha = 1$.

A natural way to accomplish this is to normalize left eigenvectors such that the weighted 2-norm

$$||\mathbf{w}|| \equiv \sqrt{\mathbf{w}^T \Pi^{-1} \mathbf{w}} \tag{S11}$$

is 1 for each left eigenvector, and to normalize right eigenvectors such that the weighted 2-norm

$$||\mathbf{v}|| \equiv \sqrt{\mathbf{v}^T \Pi \mathbf{v}} \tag{S12}$$

is 1 for each right eigenvector. Then the orthogonality condition Eq. (S10) is satisfied and so is (S9) with $\alpha = 1$. This approach has the nice side effect that the first right eigenvector $\mathbf{v}_1 = \mathbf{1}$ is a vector of 1s, and the elements of the first right eigenvector $\mathbf{w}_1 = \boldsymbol{\pi}$ sum to 1.[1]

[1] We also note in passing that Eqs. (S11) and (S12) are dual norms. If $||\mathbf{v}||$ is a norm for vector $\mathbf{v}$, the dual norm of $\mathbf{w}$ is the least upper bound of $\mathbf{w}^T\mathbf{v}$ for all $\mathbf{v}$ such that $||\mathbf{v}|| \leq 1$. To see that these are dual norms, recall the Cauchy-Schwarz inequality $|\mathbf{a}^T\mathbf{b}| \leq ||\mathbf{a}||_2 ||\mathbf{b}||_2$. Making the replacements $\mathbf{a} = \Pi^{-1/2}\mathbf{w}$ and $\mathbf{b} = \Pi^{1/2}\mathbf{v}$ lets us write this in terms of the norms Eqs. (S11)-(S12):

$$|\mathbf{w}^T\mathbf{v}| \leq ||\Pi^{-1/2}\mathbf{w}||_2 \, ||\Pi^{1/2}\mathbf{v}||_2 = \sqrt{\mathbf{w}^T\Pi^{-1}\mathbf{w}} \, \sqrt{\mathbf{v}^T\Pi\mathbf{v}} = ||\mathbf{w}|| \, ||\mathbf{v}||. \tag{S13}$$

As $|\mathbf{w}^T\mathbf{v}|$ reaches its largest value when $||\mathbf{v}|| = 1$, the least upper bound of $|\mathbf{w}^T\mathbf{v}|$ is $\sqrt{\mathbf{w}^T\Pi^{-1}\mathbf{w}}$, Eq. (S11).

S1.4 The A coordinate

In this section we focus on the first term in Eq. (S7). We show that the $A(t)$ factor in Eqs. (S3) or (S6) can be understood as a weighted average of elements of $\mathbf{R}(t)$. To see why, first consider the simpler model $\mathbf{R}(t) = P(t)\mathbf{R}(0)$ in which $A(t)$ is fixed at 1. Left-multiplying this model by the first left eigenvector $\mathbf{w}_1 = \boldsymbol{\pi}$, and exploiting the fact that $\boldsymbol{\pi}$ is the stationary state of the matrix $P(t) = e^{-L_\Phi(t/\tau)}$, we see that

$$\bar{R}(t) = \boldsymbol{\pi}^T\mathbf{R}(t) = \boldsymbol{\pi}^T\mathbf{R}(0) = \bar{R}(0). \tag{S14}$$

The left and right sides are weighted averages of elements of $\mathbf{R}$ with weights given by the elements of $\boldsymbol{\pi}$. This calculation shows that this particular weighted average is unaffected by the multiplication of $\mathbf{R}(0)$ by $P(t)$. This is a well-known fact of consensus dynamics models, which $\mathbf{R}(t) = P(t)\mathbf{R}(0)$ is an instance of: The dynamics preserves the average $\bar{R}(0) = \boldsymbol{\pi}^T\mathbf{R}(0)$ of the initial condition $\mathbf{R}(0)$ (e.g. [1]).

Next, letting $A(t)$ vary, it is clear that the average of the elements of $\mathbf{R}$ can now change over time. Left-multiplying Eq. (S3) by $\boldsymbol{\pi}$, after rearrangement we have

$$A(t) = \frac{\boldsymbol{\pi}^T\mathbf{R}(t)}{\boldsymbol{\pi}^T\mathbf{R}(0)} = \frac{\bar{R}(t)}{\bar{R}(0)}. \tag{S15}$$

This is the factor by which the average of $\mathbf{R}$ has changed from the time 0 to time t. Note that the time $t = 0$ is just a reference period in this ratio, with no special significance, and that $\bar{R}(0)$ is merely a reference value. Taking this value to be 1, we have simply

$$A(t) = \boldsymbol{\pi}^T\mathbf{R}(t) = \bar{R}(t). \tag{S16}$$

The same result can also be obtained from the eigen-expansion Eq. (S6). Left-multiplying by $\mathbf{w}_1 = \boldsymbol{\pi}_1$, and exploiting the biorthogonality of left and right eigenvectors, we have $\boldsymbol{\pi}^T\mathbf{R}(t) = A(t)c_1(0)\boldsymbol{\pi}^T\mathbf{1} = A(t)c_1(0)$, where $\boldsymbol{\pi}^T\mathbf{1} = 1$ because the elements of $\boldsymbol{\pi}$ sum to 1. The $c_1(0)$ coefficient plays a redundant role with $A(t)$, and so we set it equal to 1, leaving us with Eq. (S16).

S1.5 Review of the PCI and ECI

Here we give an expanded review of the calculations that yield the conventional PCI and ECI. This aids our discussion of the PCI* and ECI* in the next section. The material in this section is not novel and is treated in [2, 3, 4, 5].

Eigenvector definition of PCI and ECI. Let n_c be the number of countries and n_p the number of products. Let R be the $n_c \times n_p$ matrix R of RCAs, and let M be a binarized version of this matrix, with elements $M_{cp} = 1$ wherever $R_{cp} \geq 1$ and 0 otherwise. M may be viewed as an adjacency matrix for a bipartite network connecting countries to products that they export at a significant ability level. Define the diversity of country c's activity basket as the cth row sum of this matrix, denoted as d_c:

$$d_c \equiv \sum_p M_{cp}. \tag{S17}$$

Similarly, define the ubiquity of activity p as the pth column sum, denoted as u_p:

$$u_p \equiv \sum_c M_{cp}. \tag{S18}$$

Let D and U be the diagonal matrices formed from countries' diversities and products' ubiquities, respectively. One may use the products of $D^{-1}M$ and $U^{-1}M^T$, in different orders, to define two square matrices that eliminate either the country nodes or the product nodes from the bipartite network. Multiplying these matrices in one order collapses the country dimension, leading to the $n_p \times n_p$ row-normalized stochastic matrix

$$\mathcal{P} \equiv U^{-1}(M^T D^{-1} M). \tag{S19}$$

The second right eigenvector of $\mathcal{P}$ defines the unstandardized PCIs, **PCI**. Similarly, collapsing the product dimension leads to the $n_c \times n_c$ row-normalized stochastic matrix

$$\mathcal{C} \equiv D^{-1}(MU^{-1}M^T). \tag{S20}$$

The second right eigenvector of $\mathcal{C}$ defines the unstandardized ECIs, **ECI**. Often the PCIs and ECIs are standardized to have zero mean and unit variance across products or countries [4], but here we work with the unstandardized vectors, and in fact our theory implies that standardizing removes information, as we discuss in the main text.

Alternate expression for the ECI. As observed in Mealy et al. [5], the ECI of a country can also be computed as an average of the PCIs for the actvities in which the country has an RCA greater than 1. This observation is useful here, because this average formulation of the conventional ECI can be easily compared with a similar average formulation of the ECI*, which we derive in the next section.

To see that the ECI is an average of PCIs, note that $\mathcal{P}$'s eigenvectors by definition satisfy $U^{-1}(M^T D^{-1} M)\tilde{\mathbf{v}}_\mu = \alpha_\mu \tilde{\mathbf{v}}_\mu$ where α_μ is an eigenvalue. Left-multiplying this by $D^{-1}M$ gives $(D^{-1}MU^{-1}M^T)(D^{-1}M)\tilde{\mathbf{v}}_\mu = \alpha_\mu(D^{-1}M)\tilde{\mathbf{v}}_\mu$, or

$$\mathcal{C}(D^{-1}M)\tilde{\mathbf{v}}_\mu = \alpha_\mu(D^{-1}M)\tilde{\mathbf{v}}_\mu. \tag{S21}$$

We see that $(D^{-1}M)\tilde{\mathbf{v}}_\mu$ is an eigenvector of $\mathcal{C}$, and in addition, we see that the eigenvalue is α_μ, i.e. $\mathcal{C}$ shares the eigenvalues of $\mathcal{P}$. In particular, if $\mathcal{C}$'s second eigenvalue α_2 is distinct, then the associated eigenvector **ECI** is unique up to a scalar multiple, and must correspond to the vector $D^{-1}M\,\tilde{\mathbf{v}}_2$. But the elements of $\mathcal{P}$'s second eigenvector by definition are the PCIs, $\tilde{\mathbf{v}}_2 = \mathbf{PCI}$, and thus

$$\mathbf{ECI} = D^{-1}M\,\mathbf{PCI}, \tag{S22}$$

or in index form

$$\mathrm{ECI}_c = \sum_p \left(\frac{M_{cp}}{\sum_{p'} M_{cp'}} \right) \mathrm{PCI}_p. \tag{S23}$$

S1.6 The PCI* and ECI*

In this section we focus on the second term in Eq. (S7). We discuss L_Φ's second eigenvector $\mathbf{v}_2$, which we call $\mathbf{PCI}^*$, and how it is related to the conventional PCI vector $\mathbf{PCI}$. We then discuss how the coordinate $b(t)$, which we call ECI*, is related to the conventional ECI. This section focuses on theoretical relationships, while Section S2.2 examines these relationships empirically.

The PCI* The second eigenvector of the Laplacian matrix of the dynamical model, $\mathbf{v}_2$, is of special interest because it captures the pattern of compositional change with the longest time scale. Activities that take an especially long time to shift out of have negative elements in this eigenvector, and activities that take an especially long time to reach have positive elements. This mode of change represents a shift in emphasis from some activities to others, as determined by the signs of elements of the second eigenvector $\mathbf{v}_2$, thus giving this vector a meaning as a pattern of economic transformation. We now discuss further how the elements of $\mathbf{v}_2$ are closely related to the complexity metrics known as the Product Complexity Indices. To highlight its relation to the PCIs, we refer to this eigenvector as $\mathbf{PCI}^*$:

$$\mathbf{PCI}^* \equiv \mathbf{v}_2. \tag{S24}$$

To see that elements of $\mathbf{v}_2$ are related to PCIs, note that the eigenvectors of our dynamical model depend on how we construct the network of activities, and in particular on how we define proximity between activities. Prior work has generated a variety of measures of proximity, each finding empirical support when used to forecast transitions in economic activities. It is not clear what proximity measure describes the network of transitions best, and our goal here is not to resolve this. Instead, we show that the structure of the second eigenvector is not sensitive to this choice, and that across a variety of empirically-supported measures of proximity (Fig. S3), the second eigenvectors are quite similar to one another, and to the vector of the conventional PCIs.

In fact, there is a particular choice of proximity matrix for which $\mathbf{v}_2$ and the vector of PCIs ($\mathbf{PCI}$) are identical up to an irrelevant factor. This happens when the proximities are taken to be

$$\Phi^P_{pp'} \equiv \sum_c \frac{M_{cp}M_{cp'}}{D_c}, \tag{S25}$$

or in matrix form $\Phi^P = M^T D^{-1} M$. If the proximities used in the dynamical model are given by $\Phi = \Phi^P$, then $\mathbf{PCI}^*$ and $\mathbf{PCI}$ will be the second eigenvectors of the same eigenvector equation. To see this, note that $\mathbf{PCI}^*$ is by definition an eigenvector of $L_\Phi = I - \tilde{\Phi}$, and consequently an eigenvector of $\tilde{\Phi}$. Finally, $\tilde{\Phi} \equiv \mathsf{Diag}(\Phi\mathbf{1})^{-1}\Phi$ equals the matrix $\mathcal{P}$, whose second eigenvector defines the conventional PCIs:

$$\begin{aligned}
\tilde{\Phi} &\equiv \mathsf{Diag}(\Phi^P\mathbf{1})^{-1}\Phi^P \\
&= \mathsf{Diag}(M^T D^{-1} M\mathbf{1})^{-1} M^T D^{-1} M \\
&= \mathsf{Diag}(M^T D^{-1}\mathbf{d})^{-1} M^T D^{-1} M \\
&= \mathsf{Diag}(M^T\mathbf{1})^{-1} M^T D^{-1} M \\
&= \mathsf{Diag}(\mathbf{u})^{-1} M^T D^{-1} M \\
&= U^{-1} M^T D^{-1} M \\
&= \mathcal{P}.
\end{aligned} \tag{S26}$$

The close relation between the second eigenvector of our dynamical model, $\mathbf{PCI}^*$, and $\mathbf{PCI}$ carries over beyond this special case. We will show this in Section S2.2, where we demonstrate empirically that the second eigenvectors of existing proximity measures resemble one another and $\mathbf{PCI}$.

The ECI* As noted earlier, the coefficients of the eigen-expansion are obtained by exploiting the biorthogonality of the left eigenvectors $\mathbf{w}_\mu$ with the right eigenvectors $\mathbf{v}_\mu$. This property allows the coefficients of the expansion to be computed according to Eq. (S5),

$$c_\mu(t) = \mathbf{w}_\mu^\dagger \mathbf{R}(t).$$

Examining Eq. (S7), the first coefficient satisfies

$$A(t) = c_1(t) = \mathbf{w}_1^\dagger \mathbf{R}(t), \tag{S27}$$

and the second coefficient satisfies

$$A(t)b(t) = c_2(t) = \mathbf{w}_2^\dagger \mathbf{R}(t). \tag{S28}$$

It follows that the coordinate $b(t)$ is the ratio of these coefficients:

$$b(t) = \frac{c_2(t)}{c_1(t)} = \frac{\mathbf{w}_2^\dagger \mathbf{R}(t)}{\mathbf{w}_1^\dagger \mathbf{R}(t)}. \tag{S29}$$

This coordinate and the activity basket $\mathbf{R}(t)$ vary across countries. Let us now account for this in our notation, which will help us to compare the resulting formula for ECI* with the one for the ECI. We subscript the b coordinate with c, and now work with the full $n_c \times n_p$ matrix of RCAs $R = [R_{cp}]$. In index form, the $b(t)$ coordinate for country c is then

$$\begin{aligned} b_c(t) &= \frac{[R(t)\mathbf{w}_2]_c}{[R(t)\mathbf{w}_1]_c} \\ &= \frac{\sum_p R_{cp}(t) w_{p2}}{\sum_p R_{cp}(t) w_{p1}}. \end{aligned} \tag{S30}$$

We now exploit three results that we have collected to express the b coordinate in a different form. First, in the denominator, we have noted several times that the first left eigenvector is equal to the vector $\boldsymbol{\pi}$, i.e. $w_{p1} = \pi_p$. Second, in the numerator, we use the relationship that we derived earlier between the left and right eigenvectors of L_Φ, Eq. (S9), and therefore $w_{p2} = \pi_2 v_{p2}$. These changes give

$$b_c(t) = \frac{\sum_p R_{cp}(t) \pi_p v_{p2}}{\sum_p R_{cp}(t) \pi_p}. \tag{S31}$$

Finally, we insert the definition of PCI*: $\mathrm{PCI}_p^* \equiv v_{p2}$. The resulting expression has a structure that closely resembles that of Eq. (S23): Taking $\mathrm{ECI}_c^*(t) \equiv b_c(t)$, we have

$$\mathrm{ECI}_c^*(t) = \sum_p \left(\frac{R_{cp}(t)\pi_p}{\sum_{p'} R_{cp'}(t)\pi_{p'}} \right) \mathrm{PCI}_p^*. \tag{S32}$$

Like the conventional ECI Eq. (S23), the ECI* is an average of PCI*s. The conventional ECI uses uniform weights for all of the nonzero elements in the binarized matrix M. The ECI*, in contrast, uses non-uniform weights based on the elements of the unbinarized matrix R, and adjusted by the ergodic probabilities $\boldsymbol{\pi}$.

We tie the definition of PCI* and ECI* to the Laplacian L_Φ that governs the dynamics in our diversification model. This Laplacian could be constructed using different practical measures of proximity, such as Φ^P (Eq. (S25)), for example, or $\Phi^M_{pp'} \equiv \min\{(M^T M U^{-1})_{pp'}, (U^{-1} M^T M)_{pp'}\}$, the proximity matrix introduced in Ref. [6] to define the Product Space. These different proximity measures represent different guesses about underlying relationships between products, and the propensity for particular product transitions to take place. An important point is that proximity measures that yield better predictions of these product transitions would, in principle, yield better calculations of $\mathbf{PCI}^*$ and ECI* that are more informative of countries' development pathways, offering a path of refinement for these measures.

The case $\Phi = \Phi^P$. As mentioned earlier, the special case $\Phi = \Phi^P$ is notable because $\mathbf{PCI}^*$ coincides with **PCI**. We now discuss reasons why this special case is also interesting for the ECI*. First, we note that in this case L_Φ's first left eigenvector $\mathbf{w}_1 = \boldsymbol{\pi}$ is proportional to the vector of products' ubiquities (Eq. (S18)):

$$\boldsymbol{\pi} \propto \mathbf{u}. \tag{S33}$$

To see this, observe that $L_\Phi = I - \tilde{\Phi}$ and $\tilde{\Phi}$ have the same eigenvectors, and that $\mathbf{u}$ is a left eigenvector of $\tilde{\Phi} = U^{-1}M^T D^{-1} M$ (Eq. (S26)) with eigenvalue 1:

$$\mathbf{u}^T \mathcal{P} = \mathbf{u}^T U^{-1}(M^T D^{-1} M) = \mathbf{1}^T M^T D^{-1} M = \mathbf{d}^T D^{-1} M = \mathbf{1}^T M = \mathbf{u}^T. \tag{S34}$$

Taking $\boldsymbol{\pi} \propto \mathbf{u}$, Eq. (S32) can then be written

$$\mathrm{ECI}^*_c(t) = \sum_p \left(\frac{R_{cp}(t) u_p}{\sum_{p'} R_{cp'}(t) u_{p'}} \right) \mathrm{PCI}_p. \tag{S35}$$

In this special case, we see that the weights used by the ECI* are ubiquity-adjusted RCAs. These weights emerge as a by-product of using the proximity matrix $\Phi = \Phi^P$. They do not represent an intentional choice about how different products in the export basket should be weighted. However, such a ubiquity adjustment could nevertheless be beneficial from a measurement standpoint, because it puts more weight on products for which high RCAs are harder to achieve and therefore more exceptional. For example, many countries in the world produce textiles, but only a few produce ostrich eggs. As a result, almost all ostrich egg producers have high RCAs in this product, but to achieve a high RCA in textiles a country must outcompete many others. In general, achieving high RCAs will be more difficult in ubiquitous than in nonubiquitous products. It would therefore be reasonable to regard high values of R_{cp} as more impressive the more ubiquitous product p is. The conventional ECI not only thresholds RCAs in weighting products in a country's activity basket, but also makes no adjustment for how competitive each product's export market is. In contrast, the ECI* uses unthresholded RCAs, and in the special case $\Phi = \Phi^P$, it emphasizes the RCAs in ubiquitous products. We do not offer a more general interpretation of the ergodic weights π_p that applies outside the special case $\Phi = \Phi^P$, but exploring such issues could reveal interesting aspects of measurement in those contexts as well.

S1.7 Dimensionality reduction

We can relate the theory here to recent works that study the geography of economic activity and complexity metrics through the lens of dimensionality reduction. For instance, Mealy et al. (2019) [5] point out that the conventional ECI can be seen as a dimensionality reduction tool, being equivalent to a spectral clustering algorithm that orders countries along an axis in a way that tries to preserve the proximities between baskets of activities. Similarly, Sciarra et al. (2020) [7] put forward a complexity index that reduces the dimensionality of economic activity data by combining information from two eigenvectors of a country-country similarity graph. Brummitt et al. (2020) [8], in turn, apply machine learning methods to extract dimensions that characterize variation in historical export baskets across countries and time, finding axes that strongly correlate with the PCIs. In the works above, dimensionality reduction involves strategies to summarize complex data. In contrast, in exploring our dynamical theory, we used the tool of eigenmode decomposition. Eigenmode decomposition is not focused on discovering structure in data per se, but on describing a dynamical process. Nevertheless, it is well understood that these approaches are related, and here we briefly summarize the relationship.

First, the outcome of eigenmode decomposition is very similar to dimensionality reduction techniques: It returns dimensions (the eigenvectors $\mathbf{v}_\mu$) in which one can describe high-dimensional data (the matrix of activity baskets $\mathbf{R}(t)$), along with the coordinate projections of each data point onto these axes (the coefficients $c_\mu(t)$). The two approaches differ in how directions in the vector space are determined and interpreted. Typically, dimensionality reduction techniques aim to identify directions that capture large amounts of variation in data, while eigenmodes characterize coherent dynamical patterns. This difference is closely related to the distinction between proper orthogonal decomposition (POD) and dynamical mode decomposition (DMD) in applications of statistical learning to dynamical systems (e.g. [9]): Whereas POD modes optimally reconstruct a data set of snapshots of a system over time, DMD modes identify coherent structures in the dynamics.

The eigenmode decomposition not only generates a representation of data in a vector space, but shows that the model predicts that countries' activity baskets will converge to a lower-dimensional subspace that corresponds to the modes with the largest time scales. This is a well-known outcome of the interaction of structure and dynamics in networks [10, 1]. Heterogeneity in the time scales of different eigenmodes leads countries' activity baskets to collapse onto a low-dimensional subspace of possible activity baskets as differences between countries shrink first among quickly-decaying dimensions, leaving activity baskets scattered primarily along slowly-decaying dimensions. A simple example is given in Fig. S1, which depicts the phase portrait of the model if we assume that the characteristic time scale of the 2nd eigenmode is four times that of the 3rd, $\tau_2/\tau_3 = 4$. Phase lines show the path of a country that obeys the model, and show that variation along the 3rd eigenvector shrinks much more quickly than variation along the 2nd eigenvector. This illustrates how the theory here can be viewed through a dimensionality reduction lens, and indeed encourages such a perspective.

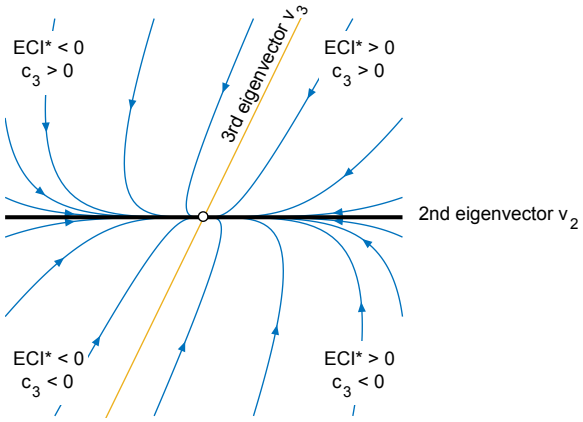


Figure S1: Illustration of model's dynamics. We display the phase portrait in the 2D plane corresponding to the 2nd and 3rd eigenmodes in a case where $\tau_2/\tau_3 = 4$. Different points in the plane correspond to different coordinate projections for an activity basket $\mathbf{R}(t)$ onto the 2nd and 3rd eigenvectors $\mathbf{v}_2$ and $\mathbf{v}_3$. We present a skewed coordinate system as a reminder that these vectors in general are non-orthogonal, $\mathbf{v}_2 \cdot \mathbf{v}_3 \neq 0$. The four quadrants are labelled by whether the projection onto each eigenvector is positive or negative. Blue lines show the paths of a country that obeys the model depending on its starting position. The model has a stable fixed point where the projections onto both eigenvectors are zero. Because the speed of convergence is much faster along the 3rd eigenvector, however, variation along this direction quickly shrinks, while variation along the 2nd eigenvector is sustained for a longer time.

S1.8 Possible extensions to structural change coordinates

Because our framework assigns different interpretations to the $\mathbf{PCI}^*$ and ECI* (as a direction of structural change, and an economy's coordinate along this direction, respectively) than usually given to the conventional PCI and ECI, our results suggest several new possibilities for using our measures to quantify structural change. We briefly comment on these possibilities here. First, note that the conventional product and economic complexity indices are typically standardized to have zero mean and unit variance, with the expectation that only their ranking is significant. However, our theory assigns meaning to the signs and magnitudes of these quantities: Elements of $\mathbf{v}_2$ ($\mathbf{PCI}^*$) convey which products are vacated or reached, and how large these compositional changes are as structural transformation occurs. Our theory therefore suggests that standardizing these quantities destroys useful information, and in some contexts should be avoided. Second, $b(t)$ (ECI*) also has cardinal meaning since it is a coordinate, and so standardizing ECIs may also destroy useful information for the same reason. In fact, the interpretation of the ECI* as a coordinate raises the possibility of tracking countries' movement over time in absolute terms, including the extent to which countries are converging towards rich-country activity baskets, something which is not possible with a rank measure. Third, by grounding these measures in the dynamics of the development process, improving the description of this process (e.g. improving how proximities Φ are estimated) should in turn feed into improved estimates of $\mathbf{PCI}^*$ and ECI*, creating a path for refinement. Finally, because our coordinates depend on the network of products Φ, our results raise the question of what happens as this network structure changes over time. Proximities between economic activities may evolve as the technologies underpinning these activities change, and this in turn would shift the principal long-term direction of activity change. The framework here could accommodate such shifts in the same way index numbers do in other settings, allowing change in a region's $b(t)$ (ECI*) coordinate to be decomposed into movement along the $\mathbf{v}_2$ ($\mathbf{PCI}^*$) direction and movement of the direction itself. Conceptually, this would represent a way to probe how much the landscape of proximities (i.e. technology) is changing, and how much different economies are converging in this landscape.

S2 Supplementary empirical results

S2.1 Transformation of RCAs

RCAs are frequently transformed nonlinearly to mitigate the influence of large values (e.g. [11, 12, 8, 13]). We do this here as well, using the function $g(R') = \log(1 + R'/R_0)$. The transformation $g(R')$ behaves linearly for small R' and logarithmically for large R', with the parameter R_0 setting the transition scale between these behaviors. Note that high values of R_0 allow large RCAs more expression while low values of R_0 suppress them (Fig. S2a). A given value of R_0 leads to a given activity vector $\mathbf{R}(t; R_0)$ and normalized activity vector $\mathbf{r}(t; R_0) \equiv \mathbf{R}(t; R_0)/A(t)$, which has a representation in the eigenvector basis

$$\mathbf{r}(t; R_0) = \sum_{\mu} b_\mu(t; R_0)\mathbf{v}_\mu = \mathbf{1} + \sum_{\mu \geq 2} b_\mu(t; R_0)\mathbf{v}_\mu. \tag{S36}$$

We choose R_0 to maximize the variance explained across countries and time by the directional vector $\mathbf{v}_2 = \mathbf{PCI}^*$, or equivalently, the variance explained by the coordinate $b_2(t; R_0) = b(t) = \mathrm{ECI}^*(t)$. Our procedure is related to principal component analysis. The difference is that a PCA *uncovers* a set of variance-maximizing directions in data, while here, the directions are given to us beforehand (the eigenvectors of the Laplacian L_Φ). Otherwise, we still ask how much variance is explained by data along the particular direction $\mathbf{v}_2 = \mathbf{PCI}^*$ for different values of R_0 and, like PCA, select a value that maximizes data variability described by this direction.

Let $\mathbf{r}_i$ be an observation of a normalized activity basket, and let $X = [\mathbf{r}_i]$ be the $n_p \times N$ data matrix of these observations across countries and years $i \in 1 \dots N$. The eigendecomposition Eq. (S36) corresponds to the matrix factorization

$$X(R_0) = VB(R_0), \tag{S37}$$

where V is the $n_p \times n_p$ matrix whose columns are L_Φ's right eigenvectors $\mathbf{v}_\mu$ and B is the $n_p \times N$ matrix whose ith column gives the coordinates of the ith observation in the eigenvector basis. The variances and covariances of the data in the directions of each vector in V can be computed as follows. Let $X_c = X\Theta$ be the centered (i.e. de-meaned) data where $\Theta \equiv I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$ is a centering matrix. The data covariance matrix is

$$\begin{aligned} C(R_0) &= \frac{1}{N}X_c(R_0)X_c(R_0)^T \\ &= \frac{1}{N}X(R_0)\Theta\Theta^T X(R_0)^T \\ &= \frac{1}{N}VB(R_0)\Theta\Theta^T B(R_0)^T V^T \\ &= \frac{1}{N}VB_c(R_0)B_c(R_0)^T V^T \\ &= V\tilde{C}(R_0)V^T, \end{aligned} \tag{S38}$$

where $B_c = B\Theta$ is the matrix of centered data coordinates in the V basis. The matrix $\tilde{C} = \frac{1}{N}B_cB_c^T$ is the covariance matrix of the data in the coordinate system given by the eigenvectors of the dynamical model.[2]

Let W be the $n_p \times n_p$ matrix whose columns are L_Φ's left eigenvectors $\mathbf{w}_\mu$. Exploiting the biorthogonality of left and right eigenvectors $W^\dagger V = I$, Eq. (S38) can be solved for $\tilde{C}$ (the covariances in the V basis) as

$$\tilde{C}(R_0) = W^\dagger C(R_0)W. \tag{S39}$$

In particular, the variance of activity baskets in the direction $\mathbf{v}_2 = \mathbf{PCI}^*$ is

$$\begin{aligned} \mathrm{Var}(\mathrm{ECI}^*; R_0) &= \tilde{C}_{22}(R_0) \\ &= \mathbf{w}_2^T C(R_0)\mathbf{w}_2. \end{aligned} \tag{S40}$$

[2] As further comparison with PCA, recall that in PCA the data is expressed in a basis V that diagonalizes the covariance matrix. In such a basis $\tilde{C}$ would be diagonal. Here, we are expressing the data in a pre-determined basis given by the eigenvectors of L_Φ, and $\tilde{C}$ will have non-zero off-diagonal elements (non-zero covariances in the resulting coordinates).

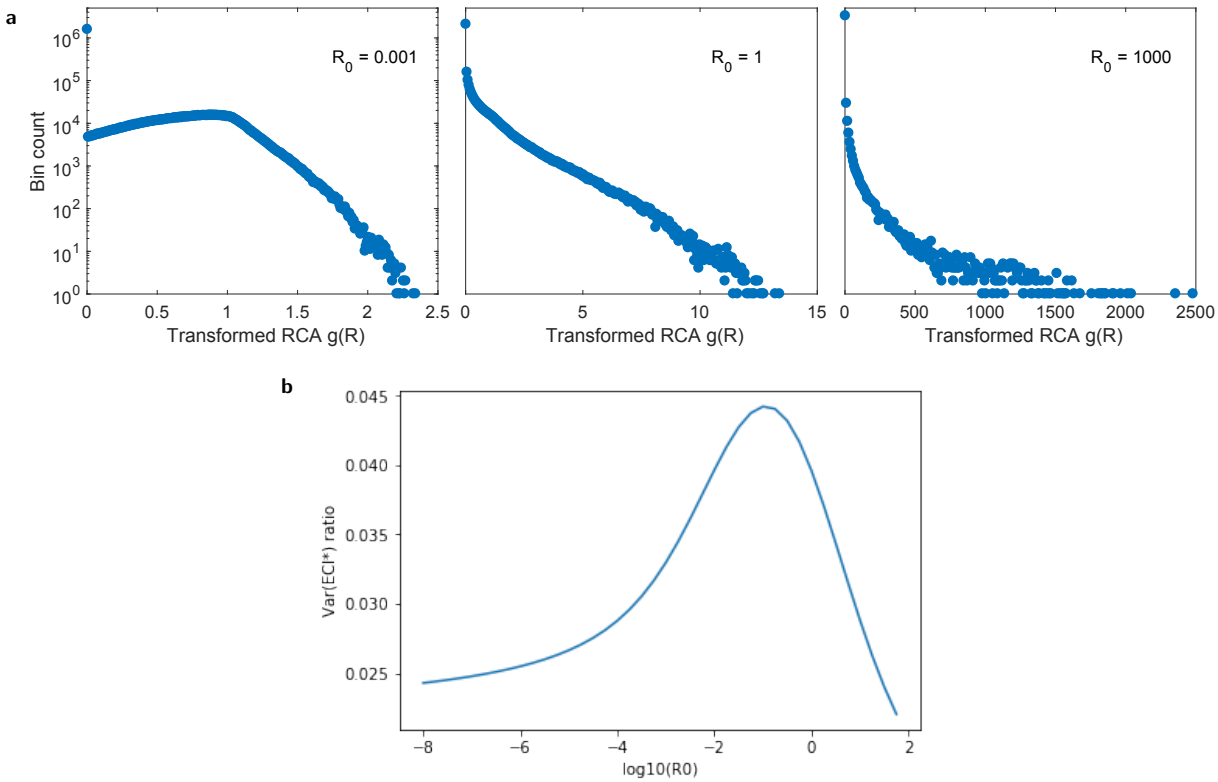


Figure S2: **a** Distribution of transformed RCAs $g(R')$ for different values of R_0. Small values of R_0 weaken the expression of the largest RCAs; large values of R_0 do the opposite. (To makes these plots easier to interpret, we divide the transformation function $g(R')$ by the constant $\log(1 + 1/R_0)$ so that a raw RCA $R' = 1$ consistently maps to a transformed RCA $R = 1$ for all values of R_0.) **b** Ratio of variance described by variation in the ECI* (Eq. (S41)) to total data variability.

The left eigenvectors were normalized with the modified 2-norm Eq. (S11), while principal components are typically normalized with a standard 2-norm. To remove the scale effect this creates we factor out the standard 2-norm from the left eigenvectors, computing the variance with $\hat{\mathbf{w}}_2 \equiv \mathbf{w}_2/||\mathbf{w}_2||_2$:

$$\widehat{\mathrm{Var}}(\mathrm{ECI}^*; R_0) = \hat{\mathbf{w}}_2^T C(R_0) \hat{\mathbf{w}}_2. \tag{S41}$$

For intuition, we could write $\widehat{\mathrm{Var}}(\mathrm{ECI}^*; R_0)$ in terms of the variances explained by principal components. Let Z be the $n_p \times n_p$ matrix whose columns are principal components. Inserting $I = ZZ^T$ above leads to

$$\widehat{\mathrm{Var}}(\mathrm{ECI}^*; R_0) = \sum_a \left(\hat{\mathbf{w}}_2^T \mathbf{z}_a(R_0)\right)^2 \sigma_a^2(R_0). \tag{S42}$$

Thus, the data variance in the direction $\mathbf{v}_2 = \mathbf{PCI}^*$ can be written as a weighted sum of the variances of the principal components, with each principal component weighted by its projection $(\hat{\mathbf{w}}_2^T \mathbf{z}_a(R_0))$ onto $\mathbf{v}_2$ in the non-orthogonal basis V. We score the variance that the coordinate $b = \mathrm{ECI}^*$ explains using the ratio of Eq. (S41) to the total data variability $\sum_a \sigma_a^2(R_0)$. We calculate this ratio for various values of R_0, finding a peak near $R_0 = 0.115$ (Fig S2b).

S2.2 Comparison of PCI*/PCI and ECI*/ECI across proximity measures

In Section S1.6 we showed how the PCI* and ECI* relate theoretically to the conventional PCI and ECI. In this section we compare these quantities empirically, showing that the PCI* is strongly correlated with the PCI and the ECI* is strongly correlated with the ECI.

The PCI* is defined by the second eigenvector $\mathbf{v}_2$ of the Laplacian matrix L_Φ. As discussed in Section S1.6, the eigenvectors of a dynamical model like the one here will depend on how one determines proximity between activities. There are several existing measures of proximity that one could use in a dynamical theory, each of which would find empirical support. Here, our goal is to show that the general structure of the second eigenvector is not particularly sensitive to this choice, and that it generically resembles the vector of conventional PCIs across a range of proximity matrices. We will similarly show that the ECI* generically resembles the conventional ECI, which is plausible, because the ECI* depends on the PCI*.

We consider three proximity matrices. The first one (Φ^P) was introduced in Section S1.6:

$$\Phi^P_{pp'} \equiv \sum_c \frac{M_{cp} M_{cp'}}{D_c}. \tag{S25}$$

We also consider the minimum conditional probability-based [6] proximity measure Φ^M that is noted in the Background of the main paper:

$$\Phi^M_{pp'} \equiv \min\{(M^T M U^{-1})_{pp'}, (U^{-1} M^T M)_{pp'}\}. \tag{S43}$$

Third, we consider a correlation-based proximity measure [13] Φ^C:

$$\Phi^C_{pp'} \equiv \frac{1}{2}(1 + \rho_{pp'}), \tag{S44}$$

where $\rho_{pp'}$ is the Pearson correlation of the RCAs for products p and p' across locations c.

Fig. S3 compares PCI*s to PCIs and ECI*s to ECIs using these three proximity measures. The first two panels show the comparison in the special case $\Phi = \Phi^P$ noted earlier. As noted in Section S1.6, when the network is constructed using the proximity matrix Φ^P, the $\mathbf{PCI}^*$ is equal to **PCI** up to a constant overall factor. Fig. S3b is the comparison of ECI* with ECI in this case, and shows the same results as Fig. 2 of the main text. Pearson and Spearman correlations are shown for each comparison, all with high values between 0.83 and 1.

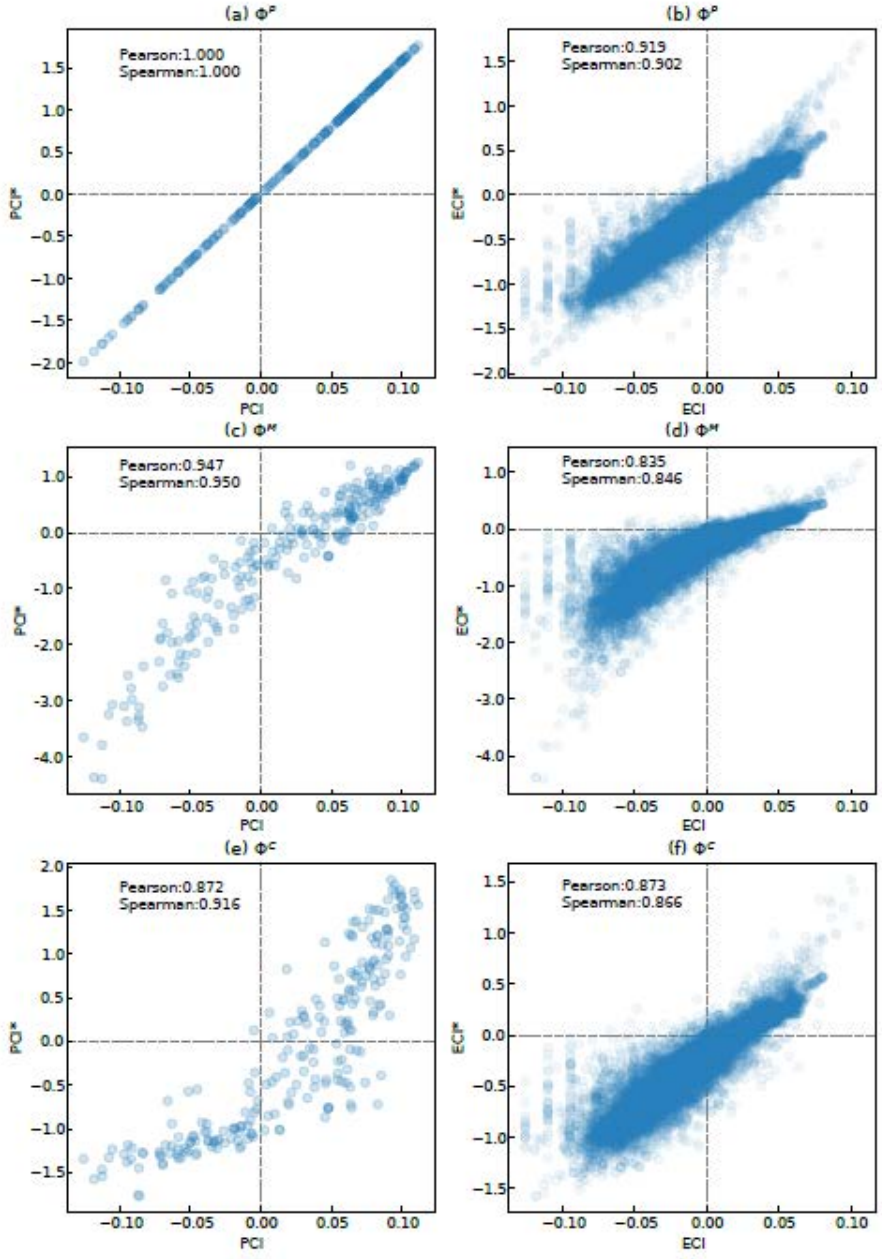


Figure S3: Comparison of PCI* with PCI and ECI* with ECI for the proximity matrices Φ^P, Φ^M, and Φ^C.

S2.3 Individual country trajectories

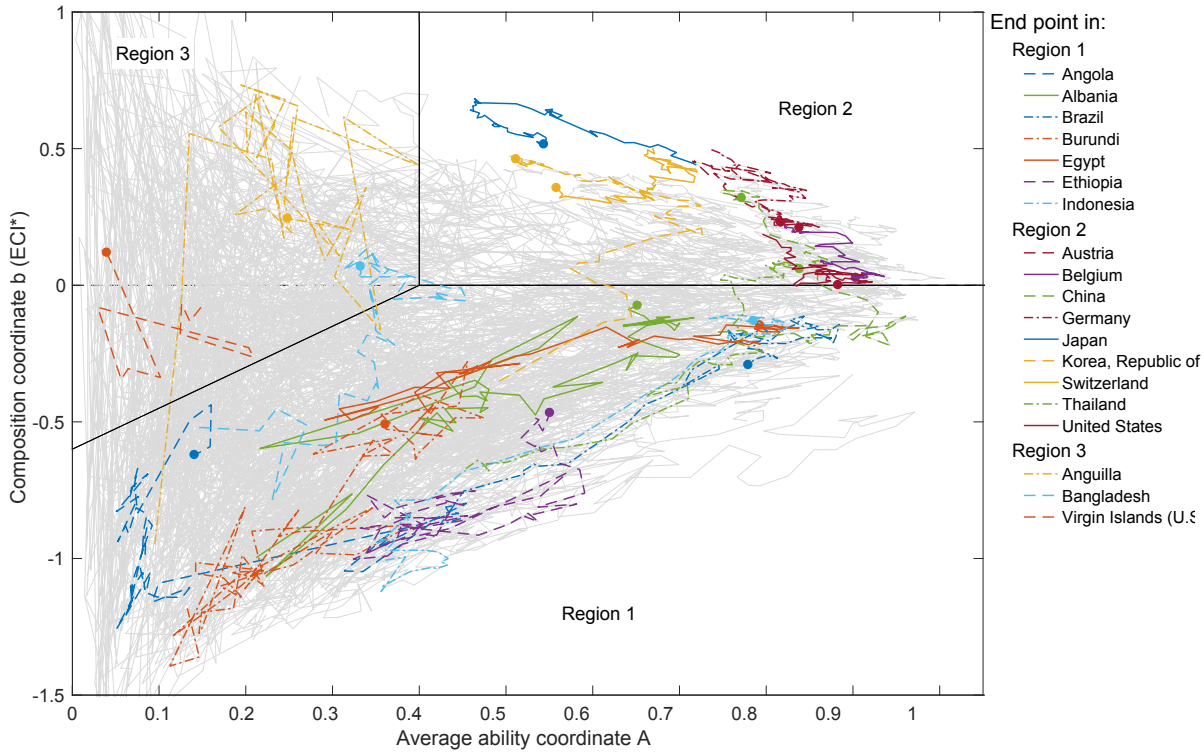


Figure S4: Trajectories of selected countries in the *A*-*b* phase plane. The labels Region 1, Region 2, and Region 3 refer to the same parts of the plane as in Figure 4e of the main paper. Round markers identify the end point of each trajectory.

S2.4 Correlations of coordinates with complexity metrics

Here we provide additional discussion about the correlations of the A and b coordinates with complexity metrics generally. We observe that existing metrics of complexity all have approximate correspondences with one or both of these coordinates. The main text and Sections S1.6 and S2.2 discussed the theoretical and empirical relationship of the ECI with the b coordinate. Another proposed complexity metric, country Fitness [14], is strongly correlated with the A coordinate (Fig. S5). This is not necessarily surprising, as Fitness is known to be strongly correlated with diversity [15], and in Fig. 2a of the main text we showed that A is also closely associated with diversity. A third metric, Production Ability [16], is also strongly correlated with the A coordinate. Finally, a recently introduced hybrid metric known as GENEPY [7] contains aspects of both the ECI and country Fitness. GENEPY is computed by first obtaining the two eigenvectors with the largest eigenvalues of a country-to-country similarity matrix. Letting $X_{c,1}$ and $X_{c,2}$ denote the elements of these eigenvectors, and letting λ_1 and λ_2 denote the eigenvalues, the GENEPY of country c is $GENEPY_c = \left(\sum_{i=1}^{2} \lambda_i X_{c,i}^2\right)^2 + 2\sum_{i=1}^{2} \lambda_i^2 X_{c,i}^2$. By design, $X_{c,1}$ is closely related to Fitness and $X_{c,2}$ is closely related to the ECI. Sciarra et al. (2020) show that, for example, when the country-country proximity matrix is $N_{cc'} \equiv \sum_p \frac{M_{cp}M_{c'p}}{d_c d_{c'}(k'_p)^2}$ where $k'_p = \sum_c M_{cp}/k_c$, then $X_{c,1}d_c = F_c$, and $X_{c,2}/\sqrt{d_c}$ is strongly correlated with ECI_c. We find empirically that the $X_{c,1}$ and $X_{c,2}$ components underlying GENEPY also have similar relationships to the A and b coordinates, i.e. $X_{c,1}d_c$ is highly correlated to A, and $X_{c,2}/\sqrt{d_c}$ is highly correlated to ECI^*.

We observe that complexity metrics and related variables fall into two groups (Fig. S5), each correlating strongly with a coordinate expected by our theory. One is a group of diversity-like quantities, including diversity d_c itself, country Fitness, Production Ability, the quantity $X_{c,1}d_c$, and the A coordinate of our theory. The other is a group of ECI composition-like quantities. These include the ECI, the quantity $X_{c,2}/\sqrt{d_c}$, and the b coordinate of our theory (ECI*). The GENEPY metric, which by design combines information associated with both groups, does not fall clearly into either one. The existence of these two groups emphasizes that, when seen through the analytical lens of this paper, existing complexity metrics emphasize different aspects of the same development process: either changes in economic diversity, or a particular pattern of compositional shifts in activities. However, as Fig. 2 in the main paper illustrated, both aspects are important characteristics of the process of development.

Sciarra et al. highlight that GENEPY merges information about countries from coordinates on two axes. Interestingly, we also arrived at a representation of country development on a 2D plane, using a very different starting point for our calculations. GENEPY is motivated by the goal of reconciling different methods for inferring complexity from data within a consistent linear algebra and networks framework. In contrast, our 2D plane arose from our analysis of a dynamical model based on the PoR. We note three differences in the resulting planes that define each of our approaches. First, as noted already above, our A and b coordinates correspond most closely to $X_{c,1}d_c$ and $X_{c,2}/\sqrt{d_c}$ rather than Sciarra et al.'s $X_{c,1}$ and $X_{c,2}$ themselves. Second, a fundamental input in the method of Sciarra et al. is a country-country proximity matrix $N_{cc'}$. The elements of the eigenvectors of this matrix then directly give countries' coordinates in the plane. In contrast, the fundamental input in the framework here is a product-product proximity matrix $\Phi_{pp'}$. Our approach comes with a two-step process: Eigenvectors derived from Φ first define directions of movement, and then country coordinates are obtained from projections onto these axes. Third, Sciarra et al. interpret their coordinates in terms of their relationship to Fitness and the ECI. Here, we exploit the fact that, by analyzing the changes in the world economy captured by movement in the directions $\mathbf{v}_1$ and $\mathbf{v}_2$, we can give economic meaning to the coordinates A and b (ECI*) that define our 2D plane, as capturing changes in average ability or shifts in activity composition.

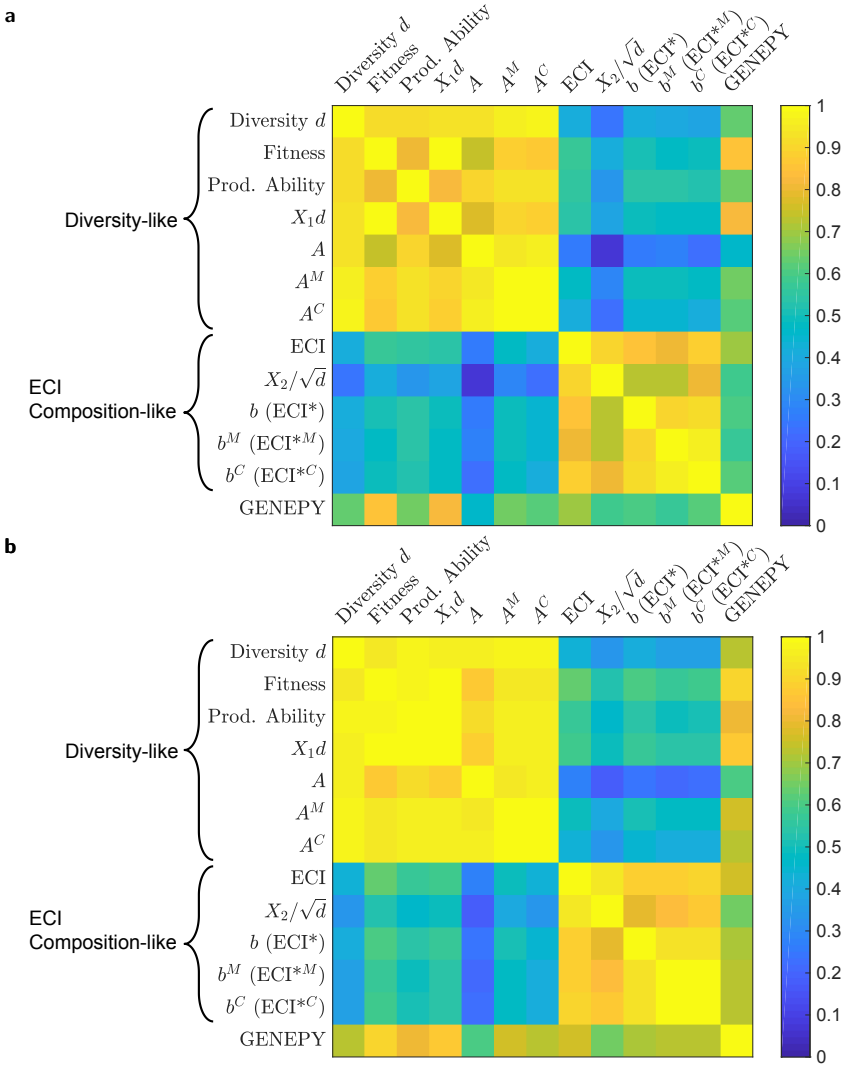


Figure S5: Correlations among our A and b (ECI*) coordinates, several proposed measures of complexity (ECI, Fitness, Production Ability, and GENEPY), and related quantities. **a** Pearson correlations. **b** Spearman rank correlations (identical to Fig. 5 in the main paper). The coordinate A is our average ability coordinate used throughout the text, defined using the proximity matrix Φ^P, while A^M and A^C refer to the same coordinate obtained from using the proximity matrices Φ^M (Eq. (S43)) and Φ^C (Eq. (S44)) respectively. Similarly ECI* is the b coordinate used throughout the text, while ECI*^{M} and ECI*^{C} are the same coordinate using the proximity matrices Φ^M and Φ^C. The quantities X_1 and X_2 are the first and second components of the GENEPY metric. All quantities were computed for the year 2016 using UN Comtrade data.

References

[1] Schaub, M. T., Delvenne, J.-C., Lambiotte, R. & Barahona, M. Structured networks and coarse-grained descriptions: A dynamical perspective. *Advances in Network Clustering and Blockmodeling* 333–361 (2019).

[2] Caldarelli, G. *et al.* A network analysis of countries' export flows: Firm grounds for the building blocks of the economy. *PloS one* **7**, e47278 (2012).

[3] Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G. & Pietronero, L. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one* **8**, e70726 (2013).

[4] Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M. & Simoes, A. *The atlas of economic complexity: Mapping paths to prosperity* (MIT Press, 2014).

[5] Mealy, P., Farmer, J. D. & Teytelboym, A. Interpreting economic complexity. *Science Advances* **5**, eaau1705 (2019).

[6] Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317** (2007).

[7] Sciarra, C., Chiarotti, G., Ridolfi, L. & Laio, F. Reconciling contrasting views on economic complexity. *Nature Communications* **11**, 1–10 (2020).

[8] Brummitt, C. D., Gómez-Liévano, A., Hausmann, R. & Bonds, M. H. Machine-learned patterns suggest that diversification drives economic development. *Journal of the Royal Society Interface* **17**, 20190283 (2020).

[9] Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L. & Kutz, J. N. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics* **1**, 391 (2014).

[10] Simon, H. A. & Ando, A. Aggregation of variables in dynamic systems. *Econometrica: Journal of the Econometric Society* 111–138 (1961).

[11] Hoen, A. R. & Oosterhaven, J. On the measurement of comparative advantage. *The Annals of Regional Science* **40**, 677–691 (2006).

[12] Elekes, Z., Boschma, R. & Lengyel, B. Foreign-owned firms as agents of structural change in regions. *Regional Studies* **53**, 1603–1613 (2019).

[13] Hausmann, R., Stock, D. P. & Yıldırım, M. A. Implied comparative advantage. *Research Policy* 104143 (2021).

[14] Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Scientific Reports* **2**, 1–7 (2012).

[15] Morrison, G. *et al.* On economic complexity and the fitness of nations. *Scientific Reports* **7**, 1–11 (2017).

[16] Bustos, S. & Yıldırım, M. A. Production ability and economic growth. *Research Policy* 104153 (2020).