Institute for
New Economic Thinking
AT THE OXFORD MARTIN SCHOOL

A New Interpretation of the Economic Complexity Index

Penny Mealy, J. Doyne Farmer & Alexander Teytelboym

4th February 2018

INET Oxford Working Paper No.2018-04

Economics of Sustainability & Complexity Economics Programmes

A New Interpretation of the Economic Complexity Index

Penny Mealy* J. Doyne Farmer† Alexander Teytelboym‡

February 4, 2018

Abstract

Analysis of properties of the global trade network has generated new insights into the patterns of economic development across countries. The Economic Complexity Index (ECI), in particular, has been successful at explaining cross-country differences in GDP/capita and economic growth. The ECI aims to infer information about countries productive capabilities by making relative comparisons across countries' export baskets. However, there has been some confusion about how the ECI works: previous studies compared the ECI to the number of exports that a country has revealed comparative advantage in ('diversity') and to eigenvector centrality. We show that the ECI is, in fact, equivalent to a spectral clustering algorithm, which partitions a similarity graph into two parts. When applied to country-export data, the ECI represents a ranking of countries that places countries with similar exports close together in the ordering. More generally, the ECI is a dimension reduction tool, which gives the optimal one-dimensional ordering that minimizes the distance between nodes in a similarity graph. We discuss this new interpretation of the ECI with reference to the economic development literature. Finally, we illustrate stark differences between the ECI and diversity with two empirical examples based on regional data.

*Institute for New Economic Thinking at the Oxford Martin School, Smith School for Enterprise and the Environment, St Edmund Hall Oxford, OX2 6ED, United Kingdom.

†Institute for New Economic Thinking at the Oxford Martin School, Department of Mathematics, Christ Church College, Oxford, OX2 6ED, United Kingdom.

‡Department of Economics, St Catherine's College, Institute for New Economic Thinking at the Oxford Martin School, Oxford, OX1 3UQ, United Kingdom.

Introduction

Structural properties of the global trade network can explain differences in economic development across countries [8, 22, 3, 4, 18, 20]. One network measure, known as the Economic Complexity Index (ECI), aims to infer information about countries productive capabilities by making relative comparisons across their export baskets [8, 4]. The ECI has been successful in explaining cross-country differences in GDP/capita and in predicting economic growth. However, there has been some confusion about what the ECI is and why it explains variation in economic development.

First, the ECI has been cast in terms of a 'corrected diversity measure (where diversity is the number of exports a country has a revealed comparative advantage in; see definition in (Eq. 1) [8, 4]. Second, the ECI has been compared to an eigenvector centrality measure [14]. This paper stresses that neither of these descriptions is accurate. The ECI is actually orthogonal to diversity [10] (i.e. the dot product between ECI and diversity vectors is 0), which means that the ECI captures a feature of the global trade network that is distinct from diversity. Moreover, we highlight that eigenvector centrality is much more closely related to diversity than to the ECI.

We show that the ECI is equivalent to a classic spectral clustering algorithm [19], which partitions a similarity graph into two parts. Hence, when applied to country-export data, the ECI represents a ranking of countries that places countries with similar exports close together in the ordering and countries with dissimilar exports far apart. More generally, the ECI can be seen as a dimension reduction tool, which gives the optimal one-dimensional ordering that minimizes the distance between nodes in a similarity graph.

Our results reveal several interesting insights for economic development and motivate potential new research avenues. First, in distinguishing the ECI from diversity, we highlight that both variables play important but different roles in the development process. Put simply, diversity captures *how many* products countries are competitive in. Country diversification patterns have been found to exhibit an inverted-U shape: countries diversify early in their developmental phase, and begin specializing at higher levels of per capita income [9]. In contrast, the ECI captures *what type* of products countries are competitive in. By grouping together countries with similar exports and separating countries with dissimilar exports, the ECI sheds light on the type of production capabilities that separate high- and low-income countries and provides empirical validation of the long-standing theory of technological capabilities in development economics [11, 12, 21].

Second, by making the link between the ECI and spectral clustering precise, we open the door for further applications of these dimension reduction tools in economic development. While partitioning the country-export similarity graph into two clusters has proven to reveal a great deal about countries productive capabilities at different developmental stages, future work could readily exploit the family of clustering approaches to glean further insights from data on economic networks.

Finally, we show that our new interpretation of the ECI helps explain its potential for applications in contexts other than international trade data. We briefly present two empirical examples from forthcoming work that illustrate the difference between the ECI and diversity in regional settings. In these two settings, ECI can explain differences in economic outcomes that cannot be captured by diversity.

The Economic Complexity Index

The ECI (and its related Product Complexity Index (PCI) measure for exported products) was originally defined using an algorithm that operates on a binary country-product matrix M with elements M_{cp}, indexed by country c and product p [8]. We say that a country c has *revealed comparative advantage* or is *competitive* in product p if $M_{cp} = 1$. The revealed comparative advantage (RCA) of country c in product p is calculated using the Balassa index [2], given by

$$RCA_{cp} = \frac{x_{cp}/\sum_p x_{cp}}{\sum_c x_{cp}/\sum_c \sum_p x_{cp}}, \tag{1}$$

where x_{cp} is country c's exports of product p. Here $M_{cp} = 1$ if $RCA_{cp} > 1$ and $M_{cp} = 0$ otherwise.

Summing across the rows and columns of M gives a country's *diversity* (denoted $k_c^{(0)}$) and product *ubiquity* (denoted $k_p^{(0)}$), defined as

$$k_c^{(0)} = \sum_p M_{cp} \tag{2}$$

and

$$k_p^{(0)} = \sum_c M_{cp}. \tag{3}$$

The ECI and PCI were originally defined through an iterative, self-referential *Method of Reflections* algorithm which first calculates diversity and ubiquity and then recursively uses the information in one to 'correct' the other [8] (see Methods).

However, it can be shown [3] that the Method of Reflections is equivalent to finding the eigenvalues of a matrix $\widetilde{M}$, whose rows and columns correspond to countries and whose entries are given by

$$\widetilde{M}_{cc'} \equiv \sum_p \frac{M_{cp} M_{c'p}}{k_c^{(0)} k_p^{(0)}} = \frac{1}{k_c^{(0)}} \sum_p \frac{M_{cp} M_{c'p}}{k_p^{(0)}}. \tag{4}$$

Equivalently, we can write $\widetilde{M}$ in matrix notation

$$\widetilde{M} = D^{-1} M U^{-1} M', \tag{5}$$

where $D = I \times M \times \mathbf{1}_{(p)}$ (the diagonal matrix formed from the diversity vector), $U = I \times M' \times \mathbf{1}_{(c)}$ (the diagonal matrix formed from the ubiquity vector), $\mathbf{1}_{(i)}$ is a vector of ones of length i, and I is the identity matrix of the appropriate dimension. Since $\widetilde{M}$ is a row-stochastic matrix (which means its rows sum to 1), the leading eigenvalue is 1, and the associated leading eigenvector is constant.

The ECI is defined as the eigenvector associated with the *second-largest* eigenvalue of $\widetilde{M}$. The precise equivalence between the Method of Reflections and the eigenvector approach is discussed in the Methods section. In the **SI**, we present the analogous derivation of the PCI.

Interpreting the $\widetilde{M}$ matrix

Given the central role $\widetilde{M}$ plays in calculating the ECI, we briefly discuss what it represents. When applied to country trade data one can think of $\widetilde{M}$ as a *weighted* similarity matrix, reflecting how similar two countries' export baskets are.

Further, from Eq. (5), we can see that

$$\widetilde{M} = D^{-1}S, \tag{6}$$

where $S = MU^{-1}M'$ is a symmetric similarity matrix in which each element $S_{cc'}$ represents the products that country c has in common with country c', weighted by each the inverse of each product's ubiquity. This formulation also makes it clear that the entries of the row-stochastic $\widetilde{M}$ matrix can also be interpreted as conditional probabilities in a Markov transition matrix [8, 10].

Results

In this paper, we denote the ECI vector by $\widetilde{y}^{[2]}$ and the ECI of country c is denoted $\widetilde{y}_c^{[2]}$. We also denote the diversity by d where $d_c = k_c^{(0)}$ is the diversity of country c. Additionally, we note that the ECI is commonly standardised by subtracting the mean and dividing by the standard deviation to allow for comparisons across years [8, 4]. However, for clarity, we use the unstandardised ECI vector throughout this paper.

ECI and Diversity

The ECI measure is conceptually cast in terms of measuring country diversity and product ubiquity, and then iteratively 'correcting' a country's diversity by the ubiquity of its products. This follows from the hypothesis that prosperous countries have capabilities that allow them to competitively export a diverse range of products which few other countries are competitive in [8, 4]. However, *the ECI is orthogonal to diversity* [10] i.e. the dot product of the diversity and the ECI vectors is zero,

$$d \cdot \widetilde{y}^{[2]} = 0. \tag{7}$$

That said, when applied to the export data, the ECI and diversity are nevertheless positively correlated, as shown in Fig. 1 (Pearson $\rho = 0.64$, p-value $= 1.1 \times 10^{-15}$). Recall that unless the mean of one of the variables is zero, a zero dot product does not imply zero correlation. Neither diversity nor the (unstandardised) ECI have zero means in the data.

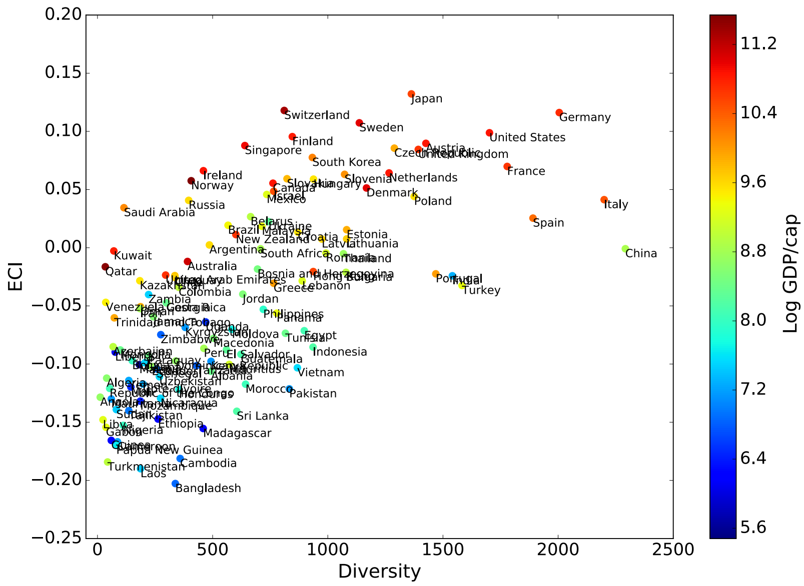

Figure 1: Country Diversity vs. ECI compared to GDP per capita. Analysis is based on HS6 COMTRADE data for the year 2013 and GDP per capita data from the World Bank.

Despite being correlated, the orthogonality of ECI and diversity suggests that they are capturing different aspects of the similarity of countries based on their export baskets. Intuitively, diversity captures *how many* exports a country is competitive in whereas the ECI captures *what type* of exports a country is competitive in.

We note that an alternative measure based on global trade data called *Fitness*

[22] is much more strongly correlated with diversity [14]. Spearman rank correlations for Fitness and diversity tend to be between 0.94 and 0.97 [13]. However, making comparisons between these the ECI and Fitness measures is beyond the scope of this paper.

ECI and eigenvector centrality

The ECI has recently been described as 'standard eigenvalue centrality algorithm' [14]. Eigenvector centrality is defined as the eigenvector corresponding to the largest eigenvalue of an adjacency matrix and this definition is standard for symmetric matrices that represent undirected networks. In the case of directed networks, the natural definition is to take the left eigenvector corresponding to the leading eigenvalue of the adjacency matrix [15, p. 178].

Therefore, the eigenvector centrality vector of $\widetilde{M}$ is the left (row) eigenvector x corresponding to the largest eigenvalue of the following eigenvalue equation

$$x\widetilde{M} = \lambda x. \tag{8}$$

Since $\widetilde{M}$ is row-stochastic, its largest eigenvalue is 1. Hence we are in interested in solutions to

$$x\widetilde{M} = x. \tag{9}$$

As the rows $\widetilde{M}$ have been normalized by diversity, it is easy to check that any vector proportional to d is a solution to Eq. (9). Since *eigenvector centrality of $\widetilde{M}$ is proportional to diversity*, it does not add anything to what we already know about $\widetilde{M}$.

From Eq. (6), we can also consider the symmetric similarity matrix $S = D\widetilde{M}$ which, when applied to country trade data, represents the similarity in countries' exports. Since S is a symmetric matrix, it represents an undirected network and is a more natural candidate for eigenvector centrality. In Fig. 2 we show that the *eigenvector centrality of S is highly correlated with diversity*. This is unsurprising. Diversity is the degree centrality of S and degree centrality is correlated with eigenvector centrality in many networks.

The fact that the ECI is defined as the eigenvector associated with the *second-largest* eigenvalue, whereas eigenvalue centrality is associated with the leading eigenvalue, makes it clear that *ECI is not eigenvector centrality.* In fact, in the export data, eigenvector centrality is much more closely correlated to diversity than to the ECI (see Fig. 2).

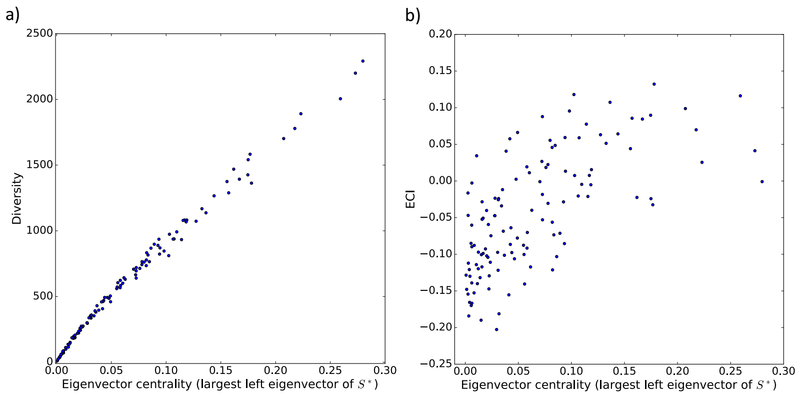

Figure 2: Panel (a) shows the eigenvector centrality of the export similarity matrix S vs country diversity (d) and Panel (b) shows the eigenvector centrality of S vs the ECI. Analysis is based on H6 COMTRADE data for the year 2013.

ECI and Spectral Clustering

We now show that the ECI is equivalent to a spectral clustering method for the problem of partitioning an undirected weighted graph, represented by an adjacency matrix S, into two components [19]. Spectral clustering is a widely used technique for community detection and dimensionality reduction and has a range of diverse applications including image recognition, web page ranking, information retrieval and RNA motif classification.

The goal of the spectral clustering approach is to minimize the sum of edge weights cutting across the graph partition, while making the size (number of nodes) of the two components relatively similar. As we discuss below, finding the exact solution to this problem is NP-hard. However, it is possible to obtain an approximate solution by minimizing the *normalized cut (Ncut) criterion* [19]. We demonstrate that the ECI is equivalent to this approximate solution.

The $Ncut$ criterion

Consider an undirected graph $G = (V, E)$ with vertices V and edges E. We allow the graph G to be weighted, with non-negative weights so the adjacency matrix entries are $S_{ij} \geq 0$ where $S_{ij} = S_{ji}$. While the export matrix is one possible example, we can consider S to be any matrix with these properties. The degree of vertex i is defined as

$$d_i = \sum_{j \in V} S_{ij}, \tag{10}$$

and the size or "volume" of a set of vertices $A \subseteq V$ can be measured as

$$vol(A) = \sum_{i \in A} d_i. \tag{11}$$

One way to partition a graph into two disjoint sets is by solving the *cut* problem. The objective is to find a partition of V into complementary sets A and $\bar{A}$ that minimize the number of links between the two sets. The *cut* problem is to find the minimum of

$$cut(A, \bar{A}) = \sum_{i \in A, j \in \bar{A}} S_{ij}. \tag{12}$$

This objective function has the undesirable property that its solution often partitions a single node from the rest of the graph. To avoid this problem, the *normalized cut* ($Ncut$) *criterion* [19] penalizes solutions that are not properly balanced. The objective is to partition the graph in such a way that each cluster contains a reasonable number of vertices. This can be achieved by minimizing the objective function

$$Ncut(A, \overline{A}) = (\frac{1}{vol(A)} + \frac{1}{vol(\overline{A})}) \sum_{i \in A, j \in \bar{A}} S_{ij}. \tag{13}$$

Let D be the diagonal degree matrix with $D_{ii} = d_i$ and $D_{i \neq j} = 0$. Then finding the minimum value of $Ncut$ is equivalent to solving the optimization problem

$$\min_A Ncut(A, \bar{A}) = \min_y \frac{y^T(D-S)y}{y^T Dy}, \tag{14}$$

subject to $y_i \in \left\{1, -vol(A)/vol(\bar{A})\right\}$ and $y^T D\mathbf{1} = 0$.

Due to the fact that y_i is restricted to one of two possible values, this is not a simple linear algebra problem, and finding the true minimum of the $Ncut$ criterion has been shown to be NP-hard [19]. However, by letting y_i take on any real value, an approximate solution can be obtained by finding the eigenvector $y^{[2]}$ corresponding to the second-smallest eigenvalue of the generalized eigenvalue equation

$$(D-S)y = \lambda Dy. \tag{15}$$

Recall that $L_S = D - S$ is called the *Laplacian* matrix of S. By making the substitution

$$y = D^{-1/2}z, \tag{16}$$

this can be rewritten as a standard eigenvalue equation

$$D^{-\frac{1}{2}}(D-S)D^{-\frac{1}{2}}z = \overline{L}_S z = \lambda z, \tag{17}$$

where $\overline{L}_S = D^{-\frac{1}{2}}(D-S)D^{-\frac{1}{2}}$ is the *normalized Laplacian* of S. Because the normalized Laplacian is a stochastic matrix its smallest eigenvalue is zero. The eigenvector $z^{[2]}$ associated with the second-smallest eigenvalue of $\overline{L}_S$, which is called the *normalized Fiedler vector*, which yields an approximate minimum to the $Ncut$ criterion [19]. Transforming back to y using Eq. (16) to solve the original problem gives the solution

$$y^{[2]} = D^{-1/2}z^{[2]}. \tag{18}$$

The solution $y^{[2]}$ provides a useful approximate solution that minimizes the normalized cut criterion and is equal to a simple transformation of the normalized Fiedler vector.

Relationship between ECI and the $Ncut$ criterion

Recall that $\widetilde{M}$ is the matrix whose eigenvector corresponding to the second-largest eigenvalue is the ECI. To see the relationship between spectral clustering and the ECI, note that the similarity matrix $S = D\widetilde{M}$ characterising country export similarity is in the same form used to minimise the normalised cut criterion. Multiplying both sides of Eq. (17) by $D^{-\frac{1}{2}}$ and re-arranging terms gives

$$D^{-1}SD^{-\frac{1}{2}}z = (1-\lambda)D^{-\frac{1}{2}}z. \tag{19}$$

Substituting $\widetilde{M} = D^{-1}S$ gives

$$\widetilde{M}D^{-\frac{1}{2}}z = (1-\lambda)D^{-\frac{1}{2}}z. \tag{20}$$

The eigenvalue equation for $\widetilde{M}$ is

$$\widetilde{M}\widetilde{y} = \widetilde{\lambda}\widetilde{y}. \tag{21}$$

Now, comparing Eqs. (20) and (21), we can see that the eigenvalues and eigenvectors of $\widetilde{M}$ are related to those of $\overline{L}_S$ by

$$\widetilde{\lambda} = 1-\lambda, \text{ and} \tag{22}$$

$$\widetilde{y} = D^{-\frac{1}{2}}z. \tag{23}$$

Thus the second-smallest eigenvalue of $\overline{L}_S$ corresponds to the second-largest eigenvalue of $\widetilde{M}$, and comparison to Eq. (18) makes it clear that the *ECI is equivalent to the spectral clustering solution of the normalized cut criterion*, i.e.

$$\widetilde{y}^{[2]} = y^{[2]} = D^{-\frac{1}{2}}z^{[2]}. \tag{24}$$

To summarize, the ECI is related to the normalized Fiedler vector by a simple transformation, in precisely the same way that the solution to the normalized cut criterion problem is related to the normalized Fiedler vector. In the **SI**, we also show how this interpretation can be applied to the PCI, and describe the precise relationship between the ECI and PCI.

The ECI partitions the country-export similarity graph

When interpreted as a clustering algorithm the ECI sorts countries into two clusters. It does this by assigning each country a real number on an interval with both positive and negative values, such that countries with similar ECI have similar exports. Countries with positive ECI are in one cluster and countries with negative ECI are in the other cluster, and the absolute value of a country's ECI measures the distance of any given country to the boundary between the clusters.

A visual representation can be seen in Fig. 3. Drawing on country trade data for 2013, we show the graph based on the similarity matrix S on the right hand side, where countries are represented as nodes and weighted links are given by $S_{cc'}$ (for visualisation purposes, we are only showing links with a weight larger than a given threshold. In this case we plot all links with a weight > 3). Here countries are coloured by their ECI, with darker shades of green representing higher complexity and darker shades of pink representing lower complexity. On the left hand side of the plot we show ECI values for each country plotted in ascending order. Countries with an ECI value greater than zero are in the green cluster and countries with a negative ECI are in the pink cluster. The ECI value (above or below zero) provides an indication of the distance from the cut.

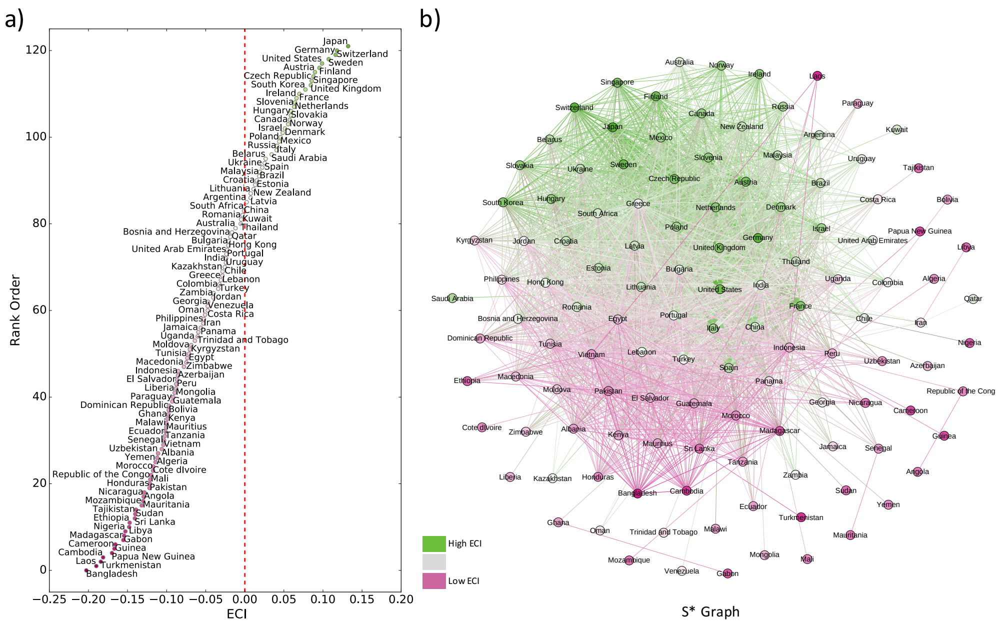

Figure 3: Visual representation of how the ECI vector partitions the graph based on similarity matrix S. The graph in Panel (b) shows country nodes coloured by their ECI value, with shades of green representing more positive ECI and shades of pink representing more negative ECI. The ECI rank ordering of countries in Panel (a) shows how countries having positive ECI fall on one side of the partition and countries having negative ECI fall on the other side of the partition.

The ECI as a dimension reduction tool

For the economic applications presented here ECI does not yield a clear separation into disjoint clusters. It nonetheless provides a useful rank ordering that places countries with similar exports near each other. To illustrate why this ordering is useful, suppose we hypothesize that there exists a relationship between the exports of a country and some quantity of interest, such as GDP per capita. To reduce the dimensionality of the problem it would be useful to find a rank ordering that places countries with similar exports close to each other. However, there are potentially $C!$ ways to order a set of C countries. For $C \approx 100$ this number is intractably large, and searching for all possible rank orderings would be impossible.

It turns out that the ECI is *the* unique way of assigning a real number to each country in order to minimize the sum of the squared distances between countries, where the distances are weighted according to the similarity matrix S [19]. While the second-smallest eigenvector $y^{[2]}$ only approximates the normalized cut criterion, it exactly minimizes

$$\frac{\sum_{ij} (y_i - y_j)^2 S_{ij}}{\sum_i y_i^2 d_i}, \tag{25}$$

subject to the constraint

$$\sum_i y_i d_i = 0. \tag{26}$$

Note how the orthogonality between ECI and diversity (Eq. 7) is hard-wired into in this minimization problem as a constraint (Eq. 26). In contrast to its application to graph partitioning, the ECI is an *exact* solution to this problem. The ECI thus provides a reduction of the high-dimensional space of countries and their exports onto a single dimension that proves to be very useful for problems such as understanding the relationship between exports and per capita GDP. In the **SI**, we discuss this result further and contrast it to more commonly used dimension reduction approaches such as principal components analysis and ordinary least squares.

The ECI can be more useful than diversity in regional settings

While the ECI has traditionally been applied to country trade data, we also provide a preview to forthcoming work which applies the ECI to data on regional employment in industries and occupations. In the context of trade data, the orthogonality between diversity and the ECI is often masked because these variables are empirically positively correlated (see Fig. 1). However, in the regional examples we present here, regional diversity and the ECI are no longer positively correlated.

Fig. 4 shows the relationship between diversity and ECI measures calculated on the basis of UK regional employment concentrations in industries and US employment concentrations in occupations. In Panel a), the ECI of UK local authorities is negatively correlated with its industry diversity. In Panel d), there is no correlation between the state ECI and occupational diversity.

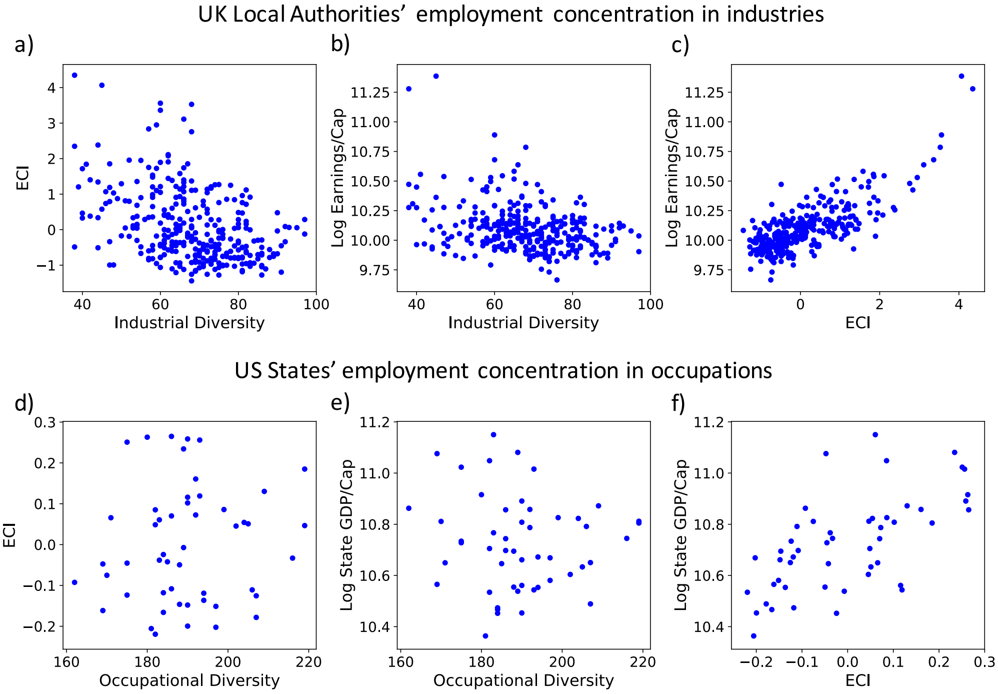

Figure 4: **Panels (a), (b), and (c)** show relationships between UK local authority industrial diversity, industrial-based ECI and log earnings per capita. Industrial employment data is sourced from the Business Register and Employment Survey for the year 2017 and 2011 work place earnings data is from the UK Office of National Statistics. **Panels (d), (e) and (f)** show relationships between US state occupational diversity, occupational-based ECI and log State GDP/cap. Occupational employment data is based on IPUMS data for the year 2010. State GDP per capita (2010) data is from the Bureau of Economic Analysis

Interestingly, in these two examples, we find that the ECI continues to be significantly positively correlated with regional per capita earnings and income (for UK local authorities, Pearson ρ between ECI and log earnings per capita $= 0.76$, p-value $= 3.9 \times 10^{-61}$, and for US States, Pearson ρ between ECI and GDP per capita $= 0.63$, p-value $= 7.46 \times 10^{-7}$). To the extent that a region's employment concentrations in industries or occupations capture information about its productive capabilities, our results suggest that the ECI is also able to shed light on the *type* of capabilities that separate rich and poor regions. Diversity, on the other hand, fails to meaningfully distinguish between high- and low-income regions in these two examples.

Discussion

In this paper we have shown that, rather than being linked to export diversity or eigenvector centrality, the ECI is equivalent to the spectral clustering solution of a normalized cut criterion [19]. Our results have some important implications for the application of these measures to the development context. In particular, by making the difference between ECI and diversity explicit, we can distinguish between the roles these measures play in the development process.

The relationship between diversification and development is well established in the economics literature. The general finding is that countries tend to follow a U-shaped pattern, where they first diversify and then begin specialising relatively late in the development process [9]. This pattern aligns with other empirical studies that have described a positive association between export diversification and economic growth, which tends to be stronger for less developed countries [1, 6, 7].

In contrast to diversity, the ECI provides additional information relevant for economic development. As we illustrate in Fig. 3, the ECI provides a rank ordering of countries in terms of how similar their exports are to each other. The fact that this ordering is useful in explaining variation in per capita GDP and predicting growth suggests that different types of exports (and by extension, productive capabilities) are associated with different growth and development outcomes. There is, indeed, empirical evidence that specializing in some products can lead to higher economic growth than specializing in others [5]. Manufacturing products in particular show strong unconditional convergence in labour productivity [16]. Within developing countries, technologically sophisticated products are more strongly associated with export and income growth [12].

We have also shown that the ECI can be thought of as a dimension reduction tool. It is the unique way to assign distances to countries such that the sum of their squared distances from each other is minimized, where the distance is measured using a weighted similarity matrix S. The ECI thus takes the complicated high dimensional space of countries and their exports and reduces it to a linear ordering, analogous to the Dewey-Decimal System for classifying books. The proven empirical success of applying this particular clustering algorithm to the country-export similarity matrix opens the door for future work to examine potential applications to the wider family of clustering and dimension reduction tools to learn further insights from the global trade network—or other data, such as country input-output tables.

Finally, we have shown two empirical examples to illustrate how in some settings, the ECI can be more useful than diversity in explaining variation in regional earnings and income per capita. Moreover, our new interpretation of the ECI suggests that, at least for these particular cases, the *type* rather than the *number* of industries and occupations concentrated in a region matters more for its economic prosperity.

Methods

The Method of Reflections and its equivalence to the $\widetilde{M}$ eigenvalue problem

Following original Method of Reflections, we can take the measures of country diversity and product ubiquity (Eqs. 2 and 3) as a starting point and then recursively calculate the average values associated with country and product nodes' neighbors from the previous iteration step as shown in Eqs. (27) and (28), yielding

$$k_c^{(N)} = \frac{1}{k_c^{(0)}} \sum_p M_{cp} k_p^{(N-1)}, \tag{27}$$

and

$$k_p^{(N)} = \frac{1}{k_p^{(0)}} \sum_c M_{cp} k_c^{(N-1)}. \tag{28}$$

As $N \to \infty$ these variables converge to constant vectors, i.e. $k_c^{(\infty)} = k$ independent of c. Originally, values associated with a moderately large value of N ($N = 18$) were considered [8]. These measures produced useful deviations from the constant vector, and were shown to have a strong positive correlation between $k_c^{(N)}$ and log GDP per capita [8].

The Method of Reflections was later reframed as an eigenvalue problem [3, 4], where it is shown that inserting Eq. (28) into Eq. (27) and rewriting gives

$$\begin{aligned} k_c^{(N)} &= \frac{1}{k_c^{(0)}} \sum_p M_{cp} \frac{1}{k_p^{(0)}} \sum_{c'} M_{c'p} k_{c'}^{(N-2)} \\ &= \sum_{c'} k_{c'}^{(N-2)} \sum_p \frac{M_{cp} M_{c'p}}{k_c^{(0)} k_p^{(0)}} \\ &= \sum_{c'} \widetilde{M}_{cc'} k_{c'}^{(N-2)}, \end{aligned} \tag{29}$$

where

$$\widetilde{M}_{cc'} \equiv \sum_p \frac{M_{cp} M_{c'p}}{k_c^{(0)} k_p^{(0)}} = \frac{1}{k_c^{(0)}} \sum_p \frac{M_{cp} M_{c'p}}{k_p^{(0)}}. \tag{30}$$

The *Economic Complexity Index* (ECI) can then be defined as the eigenvector $\widetilde{y}^{[2]}$ associated with the second-largest eigenvalue of $\widetilde{M}$. Since $\widetilde{M}$ is row-stochastic, the leading eigenvalue is one and the leading eigenvector is constant. The iterative method gives essentially equivalent results due to the fact that the eigenvector for the second-largest eigenvalue corresponds to the direction in which the system converges most slowly onto the leading eigenvector.

Note that with the earlier recursive definition, while diversity and ubiquity are given as initial conditions, they become irrelevant in the limit as $N \to \infty$. Eqs. (27) and (28) define a linear dynamical system with a stable fixed point attractor, in which the solution is independent of the initial condition. Diversity and ubiquity are relevant only because they are incorporated into the definition of the dynamical system itself.

Calculating the ECI for UK and US regional employment data

UK Local Authorities and Industries

Using data from the UK Business Register and Employment Survey (BRES), we construct a binary *region-industry matrix* W on the basis of a region rs *Location Quotient* (LQ) in industry i

$$LQ_{ri} = \frac{e_{ri} / \sum_p e_{ri}}{\sum_r e_{ri} / \sum_r \sum_i e_{ri}}, \tag{31}$$

where e_{ri} is the number of people employed in industry i in region r and $W_{ri} = 1$ if $LQ_{ri} > 1$ and $LQ_{ri} = 0$ otherwise. Note that Eq. (31) is analogous to Eq. (1). We then constructed $\widetilde{W}$ matrix from W is the same way as $\widetilde{M}$ is constructed from M (Eq. 5). Finally, we calculate the *industry-based* ECI for UK Local Authorities by finding the eigenvector associated with the second-largest eigenvalue of $\widetilde{W}$.

US States and Occupations

We apply the same methodology to calculate the *occupation-based* ECI for US states. (We also find consistent results using data on US states and industries.) Drawing on census data for the US, which is available from the Integrated Public Use Microdata Series (IPUMS) [17], we construct a *state-occupation matrix* using state's location quotient in *occupation* i. We then compute the *occupation-based* ECI for US states analogously to the industry-based ECI for UK Local Authorities.

References

[1] F. Al-Marhubi. Export diversification and growth: an empirical investigation. *Applied economics letters*, 7(9):559–562, 2000.

[2] B. Balassa. Trade liberalisation and "revealed" comparative advantage. *The Manchester School*, 33(2):99–123, 1965.

[3] G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala, and A. Tacchella. A network analysis of countries export flows: firm grounds for the building blocks of the economy. *PloS one*, 7(10):e47278, 2012.

[4] R. Hausmann, C. A. Hidalgo, S. Bustos, M. Coscia, A. Simoes, and M. A. Yildirim. *The Atlas of Economic Complexity: Mapping paths to prosperity*. MIT Press, 2014.

[5] R. Hausmann, J. Hwang, and D. Rodrik. What you export matters. *Journal of Economic Growth*, 12(1):1–25, 2007.

[6] D. Herzer and F. Nowak-Lehnmann D. What does export diversification do for growth? an econometric analysis. *Applied economics*, 38(15):1825–1838, 2006.

[7] H. Hesse. Export diversification and economic growth. *Breaking into new markets: emerging lessons for export diversification*, pages 55–80, 2009.

[8] C. A. Hidalgo and R. Hausmann. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575, 2009.

[9] J. Imbs and R. Wacziarg. Stages of diversification. *The American Economic Review*, 93(1):63–86, 2003.

[10] E. Kemp-Benedict. An interpretation and critique of the method of reflections. Mimeo, 2014.

[11] S. Lall. Technological capabilities and industrialization. *World development*, 20(2):165–186, 1992.

[12] S. Lall. The technological structure and performance of developing country manufactured exports, 1985-98. *Oxford development studies*, 28(3):337–369, 2000.

[13] M. S. Mariani, A. Vidmer, M. Medo, and Y.-C. Zhang. Measuring economic complexity of countries and products: which metric to use? *The European Physical Journal B*, 88(11):293, 2015.

[14] G. Morrison, S. V. Buldyrev, M. Imbruno, O. A. Doria Arrieta, A. Rungi, M. Riccaboni, and F. Pammolli. On economic complexity and the fitness of nations. *Scientific Reports*, 7(1):15332, 2017.

[15] M. Newman. *Networks: an introduction.* Oxford University Press, 2010.

[16] D. Rodrik. Unconditional convergence in manufacturing. *The Quarterly Journal of Economics*, 128(1):165–204, 2012.

[17] S. Ruggles, K. Grenadek, R. Goeken, J. Grover, and M. Sobek. Integrated public use microdata series: Version 7.0, 2017.

[18] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini. Randomizing bipartite networks: the case of the world trade web. *Scientific reports*, 5:10595, 2015.

[19] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[20] M. J. Straka, G. Caldarelli, and F. Saracco. Grand canonical validation of the bipartite international trade network. *Physical Review E*, 96(2):022306, 2017.

[21] J. Sutton and D. Trefler. Capabilities, wealth, and trade. *Journal of Political Economy*, 124(3):826–878, 2016.

[22] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero. A new metrics for countries' fitness and products' complexity. *Scientific reports*, 2:723, 2012.

Acknowledgements

This project was supported by Partners for the New Economy and the Oxford Martin School project on the Post-Carbon Transition. We would also like to thank Simon Angus, R. Maria Del Rio Chanona, Neave O'Cleary, Devavrat Shah, and Muhammad Yildrim for useful conversations and Ricardo Hausmann, Eric Kemp-Benedict and Luciano Pietronero for valuable comments on the original draft of this paper.