Discrimination of economic Input-Output networks using Persistent Homology

Andrew Schauf[1,*], Jae Beum Cho[2,*], Masahiko Haraguchi[3,*], and Jarrod J. Scott[4,*]

[*] *2015 Complex Systems Summer School, Santa Fe Institute*
[1] *Graduate School for Integrative Sciences and Engineering, National University of Singapore*
[2] *Department of City and Regional Planning, Cornell University*
[3] *Department of Earth and Environmental Engineering, Columbia University*
[4] *Bigelow Laboratory for Ocean Sciences*

Abstract

Viewing Input-Output (IO) tables as weighted complex networks, we investigate how certain characteristics of an economy are associated with the internal structural "shape" defined by its IO flows, as considered separately from its absolute magnitude. In this initial exploration, we examine domestic Input-Output table data from the Organisation for Economic Co-operation and Development (OECD) for 62 national economies from 7 different years spanning 1995 to 2011. By normalizing link weights so that information about absolute magnitudes is discarded, we consider only a network's "shape" as described by the relative magnitudes of flows between sectors. These normalized networks are then compared using three different similarity measures: one *topological* in the sense that it captures certain structural properties abstracted from their particular locations within the network, and others that are *"geometric"* in that they involve direct comparisons of corresponding intersectoral links at the same positions within their respective economies. Clustering analyses then indicate which aspects of an economy might be associated with the structural features captured by each of these perspectives. The topological perspective, provided here by zero-dimensional persistent homology barcodes of Input-Output networks, seems to distinguish economies of different magnitudes (as measured though the natural log of GDP and population) despite its consideration of network "shapes" as independent of their absolute sizes. This discrimination of size also coincides with differences in import and export percentages of GDP, demonstrating that the internal topological connectivity of an economy can provide indirect insight into its absolute size and the nature of its external flows. Meanwhile, "geometric" similarity measures distinguish economies in terms of a distinct set of indicators, confirming the presence of purely topological hallmarks of certain economic properties to which "geometric" perspectives are indiscriminate. Along with these preliminary observations, we discuss the potential for applying higher-dimensional persistent homology to study IO networks.

1 Introduction

1.1 Motivating questions

Many of the metrics traditionally used to characterize and compare different economies measure, in some sense, the "sizes" of certain aspects of those economies, even when considered as percent changes or *per capita* quantities. Even making one-to-one comparisons of multiple indicators or sectors may fail to capture more abstract, overarching structural similarities or differences between economies. This project grew out of a curiosity about what properties of economies might be visible not only through measurements of their sizes, but which are manifest instead in purely *structural* features which are, in some sense, independent of

size and perhaps even independent of their particular "locations" within a structure. Apart from the overall *size* of an economy, one might ask what kind of information is carried by its internal *"shape"*.

More specifically, we were guided by the following questions:

- How do the *structures* of economies – considered separately from the absolute scale of the quantities involved insofar as is possible – differ for qualitatively different types of economies?
- Can certain structural properties or features give insight into the characteristics of different economies, or help to explain and predict how they will behave under various circumstances?
- How do these structures evolve and fluctuate in response to various external processes and events? How do structures tend to change as an economy develops?
- How are the structures perturbed and how do they recover when affected by external or internal "shocks"? Do certain structural features signify different types of resilience or vulnerability?
- Can certain types of structural features – such as certain 'motifs' or patterns of interconnectivity – serve similar functions in different economies despite having different "locations" within the economy?

1.2 Input-Output networks

Most notions of economic "structure" – whether they involve the structure of relationships between different organizations, international trade patterns, supply chains, or the flow of goods and services between economic sectors – can be captured by weighted network models, where vertices corresponding to the constituent parts of the system of interest and with link weights quantifying the interactions between them. The particulars of this partition, including scale, granularity, and particular groupings of economic entities used, will of course depend on the problem at hand. Any such weighted network models of economic entities (for example "urban metabolism" networks [2],[27]) may be amenable to the types of analysis discussed here, but we begin our work by focusing on national Input-Output networks of domestic product flows [20] due to their ubiquity in economics research, their size, and the abundance and breadth of data available.

1.2.1 Input-Output tables

Input-Output (IO) tables describe the exchange of goods between the different sectors within an economy [17]. The intersectoral IO table entries of an economy partitioned into N disjoint sectors can be presented as an N-by-N matrix $\mathbf{T}$, with rows and columns corresponding to various sectors so that entries T_{ij} represent the magnitude of the directed flow from sector i to sector j. This matrix is typically asymmetric, with nonzero diagonal entries T_{ii} representing the internal flows exchanged between subentities grouped within a common sector i.

1.2.2 IO tables as weighted networks

These IO table matrices can also be interpreted as adjacency matrices of **weighted networks**, with vertices corresponding to sectors and the link weights corresponding to the magnitudes of monetary flows between those sectors. Since IO networks for different economies can be expressed in terms of the same sector definitions, there is a one-to-one correspondence between nodes and links of different IO networks. In this way, IO networks provide a natural medium in which to address the motivating questions behind this project: the objects of interest (economic sectors as vertices and intersectoral flows as links) themselves do not differ in number nor in nature – even between economies with vastly different sizes or qualitative characteristics –

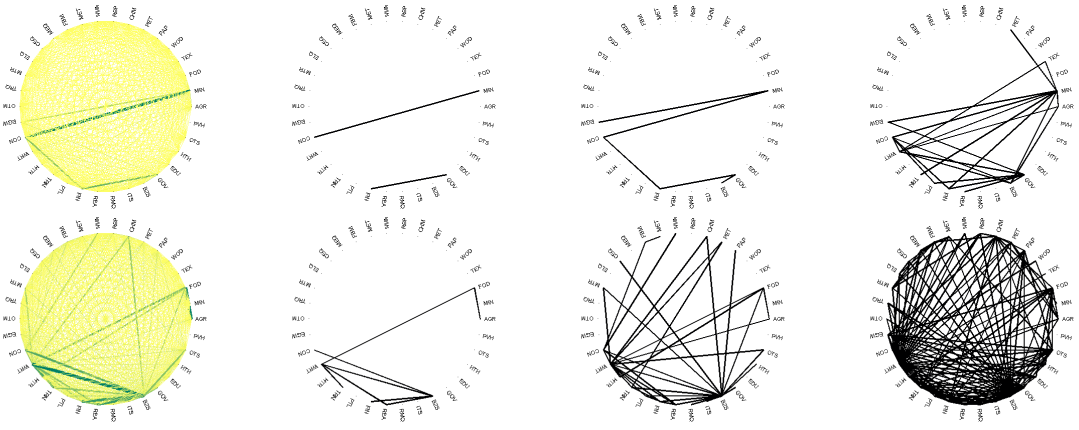

Figure 1: From left to right: Weighted network visualizations (far left) and unweighted networks for threshold values of 0.50, 0.20, and 0.05 (second from left to far right) for Bahrain (upper row) and France (lower row) IO networks from 2008. In France's IO network, more links are introduced at lower values of the threshold parameter, so that the total number of connected components of sectors more rapidly decreases towards 1.

and the important differences between the networks lie in the details of the link weights and the structural features that they collectively form. This allows the networks to be studied both *topologically* – that is, in terms of fundamental types of structural features, such as the overall number and distribution of various types of topological features, regardless of their particular locations within the structure – and *"geometrically"*, with attention to the particular sectors and links which compose these features.

1.3 Analysis of weighted IO networks

1.3.1 Analysis of IO networks by traditional network-theoretical methods

Weighted networks present challenges to many of the standard techniques typically used in the analysis of unweighted networks, in which links between pairs of vertices are either present or not (represented by adjacency matrix entries assuming values of either 1 or 0), with no notion of distinct link weights which may assume values from some continuous range of magnitudes. Many of the network-theoretical measures and techniques used to analyze unweighted networks can be generalized in more or less straightforward ways to treat weighted networks; these have even been used to study IO networks in a variety of literature [3], [28], [22], [14], [10], [1], [8], [18], [9], [11], [25].

1.3.2 Analysis of IO networks as a sequence of unweighted networks

An alternative approach to handled weighted networks is to convert the to unweighted networks by using some choice of threshold; if a certain cutoff value is chosen, and all links with weights falling below this value are discarded, the network formed by the remaining links can be studied as an unweighted network representing the essential underlying network structure (Figure 1). When there is no obviously meaningful single choice for this threshold value, a restriction to any single thresholded network is arbitrary and inherently discards potentially useful structural information. However, the entire progression of structural changes – such as the appearances or disappearances of certain topological features – that occur as one sweeps through the full range of values of this threshold parameter provides an alternative complete description of the network. This perspective shifts the focus away from properties of individual nodes or links themselves, instead viewing the network as an aggregate of topological features characterized by their "depths".

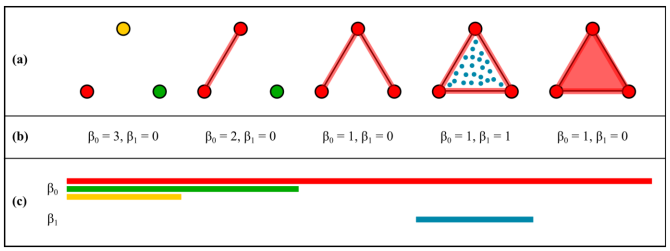

Figure 2: A simple example of a filtration of a simplicial complex, along with its "barcode", from Carstens and Horadam [7]. In the first step of the filtration, the complex comprises 3 unconnected vertices, reflected in the 3 connected components counted by β_0. In the second step, one link is introduced so that one connected component (yellow) is absorbed into another (red), thus truncating the yellow bar. In the third step, another link is introduced so that the green vertex is absorbed into the red connected component, truncating the green bar. The new link introduced in the fourth step closes the loose edges of the chain of links, forming a 1-cycle that encloses a "hole" represented by the blue bar in β_1. In the final step, a 2-simplex is introduced in the interior of this 1-cycle; this fills in the hole such that the cycle disappears and the blue bar is truncated.

1.3.3 Topological perspective: Persistent homology of weighted networks

An emerging technique which takes such a perspective is *persistent homology* (PH) [6] (see Appendix A for a more thorough introduction to the topic). The technique applies ideas from algebraic topology to identify "holes" of various types within data, and then tracks the appearances and disappearances of these features as the structure is progressively built up through some sequence of changes called a *filtration*. The record of these appearances and disappearances throughout the sequence is summarized in a *persistence diagram* or *"barcode"* which can serve as a sort of topological "fingerprint" representing the data's structure (see Figure 2). Equipped with some metric quantifying the similarity between pairs of these profiles, structures can be compared and classified in terms of their topology.

A rapidly-growing literature explores applications of persistent homology to weighted networks, developing the underlying theory and exploring various specific applications [15], [7], [24], [23]. Typically, the networks are viewed in terms of their *clique complex* or *flag complex*, in which n-cliques (complete subgraphs of n vertices) are treated as n-dimensional objects which can connect with one another to bound "holes" of dimension n. In one case, zero-dimensional persistent homology barcodes proved capable of discriminating the connectomes (correlation networks between regions of interest in brains as measured by functional MRI) representing the brains of non-ADHD-diagnosed patients from those of ADHD-diagnosed patients' brains [16].

1.3.4 Zero-dimensional versus higher-dimensional Persistent Homology

Many applications of persistent homology made so far along these lines (including [16]) have focused only on zero-dimensional homology, in effect simply counting the number of separate connected components present at each stage in the link weight parameter sweep as a relatively simple proof-of-concept. Roughly speaking, this perspective characterizes how uniformly the links weights are distributed throughout the network, a property which often has very tangible consequences for the network's operation. Although this clearly contains important information, and its interpretation is often very straightforward, an identical analysis could arguably be accomplished in a much more straightforward way by simply counting connected components using traditional graph theoretical algorithms, or maybe even other measures of link "diversity" such as the Theil index, without any explicit allusions to the formalism of algebraic topology. These proofs-of-concept are perhaps interesting, however, as first steps towards similar analyses based on higher-dimensional persistent homological features, like 1-dimensional holes ("tunnels") or 2-dimensional holes ("voids"). If

persistent homology is to prove uniquely useful for a certain application, it may be through higher-order "holes" of dimension 1 and above, which may offer more genuinely novel insights which justify the challenging conceptual learning curve and higher computational costs associated with persistent homology methods. Since the metrics initially proposed for comparing persistent homology diagrams often require a one-to-one correspondence between topological "holes" (a condition always trivially satisfied for zero-dimensional barcodes of networks with a common number of nodes, but seldom for higher-dimensional features), analysis of these higher-order features can be much more difficult to accomplish in practice. In addition to these and other technical challenges, the interpretation of higher-dimensional homology may also be less straightforward than its zero-dimensional counterparts: while the significance of a very uniformly-connected network may be easy to elucidate, precisely what are the implications of having a 1-dimensional tunnel or 2-dimensional void formed among groups of flows between sectors of an economy?

2 Methods

2.1 Scope and Input-Output data set

We begin to address the motivating questions of the project in this report by first attempting a proof-of-concept using an analysis of PH barcodes of dimension 0, here applied to 34-sector domestic Input-Output table data for 63 national economies [20]. By clustering these economies into groups based on comparisons of PH barcode profiles, and by then comparing various economic statistics averaged over the resulting clusters, we hope to gain some perspective into what the persistent homology might actually be telling us. By comparing the topological perspective to one which involves one-to-one comparisons of corresponding intersectoral flows, we hope to distinguish which kinds of economic properties depend not just on the relative magnitudes of certain sectors, but on more abstract topological properties which may be shared even by economies which superficially appear to be much different. Beyond investigating the potential of persistent-homology-based comparisons as a purely computational "black box", we ultimately hope to use the approach to gain intuition and understanding regarding any meaning that the different types of topological holes found in IO networks may carry.

Here we focus on IO tables from the Organisation for Economic Co-operation and Development (OECD) International Input-Output tables [20], [26]. This data set includes 43-sector Input-output tables for 62 countries (including estimates for aggregated "Rest of World" Input-Output data) for 7 years (1995, 2000, 2005, and 2008-2011). This allows for comparisons of IO network structure across countries and also their evolution over time. In the current project, we initially focus on IO networks composed of flows between sectors within individual countries, viewing them as self-contained structures representing the economies which can be separately normalized and compared. However, as the data set also details international flows between sectors, it also allows for future consideration of the full global IO network as well as any multi-national subsets of interest.

2.2 Converting IO tables to simple undirected IO networks

In general, IO table matrices are not symmetric, and contain nonzero entries on the diagonal representing intrasector flows. Here, we recast N-by-N directed IO tables $\mathbf{T}$ as simple, undirected network adjacency matrices A using

$$\mathbf{A} = \frac{1}{2\sum_{i,j} T_{i,j}} \left(\mathbf{T} + \mathbf{T}^T\right) \cdot \left(\mathbf{J}_N - \mathbf{I}_N\right), \tag{1}$$

where $\mathbf{J}_N$ indicates the N-by-N matrix with all 1 entries, $\mathbf{I}_N$ is the N-by-N identity matrix, and the dot indicates element-wise multiplication. That is, we take the symmetric part of the network with the self-loop edges ignored, normalized such that the sum of all entries is always equal to 1 before self-edges are

removed. The link weights $A_{i,j}$ thus give the relative importance of the total interaction between sectors i and j relative to the total magnitude of internal flows within the economy; two different IO tables $\mathbf{T}_1$ and $\mathbf{T}_2 = c\mathbf{T}_1$ representing economies of different sizes ($c \neq 1$) would nonetheless have identical *structures* as captured by their identical IO networks $\mathbf{A}$.

We note that the relative importances of the intrasectoral flows $T_{i,i}$ are included in the normalization before the diagonal is discarded such that link weights reflect the fraction of the total domestic flows which they represent. As a result, two otherwise-identical IO tables having different self-edge weights would yield networks which differed by a scale factor; in this way, the relative importance of intrasectoral flows is included indirectly in this characterization of network structure. We note that self-edges and the directedness of intersectoral flows do not affect the persistent homology, since these do not influence the clique complex constructed from the weighted network; however, in the following analyses, we also compare networks using "geometric" similarity measures which potentially *would* yield different results if applied instead to networks that still included intrasectoral flow data, or still distinguished the bidirectional flows between sectors which here are combined into single intersectoral link weights.

2.3 Computing Persistent Homology

Persistent homology diagrams for dimensions 0, 1, and 2 are computed for the normalized IO networks using *Holes* [24], a Python module for computing PH of the clique complex of weighted networks based on the JavaPlex engine [6], [21]. The outputs are used to compute barcode similiarities used in clustering analyses, and to produce the barcode visualizations displayed.

2.4 Clustering analysis

2.4.1 Network similarity via persistent homology β_0 barcodes

The zero-dimensional persistent homology barcode displays each connected component as a bar beginning at zero (since at the beginning of the filtration, no links are included and all nodes represent their own connected components) and ending at the point in the filtration at which that connected component is absorbed into another. If all links are nonzero, only a single connected component will persist until the end of the filtration. When these bars are "stacked" with more persistent components placed below less persistent components, the vertical height of the stack counts the number of connected components β_0 as a function of the weight threshold parameter w_t. The barcodes can thus be considered as monotonically decreasing functions $\beta_0(w_t)$ which descend from a common maximum value corresponding to the number of nodes in the network (at the threshold's maximum value, all nodes are disconnected) to a value of 1 (as the threshold reaches 0, the network becomes a complete graph with a single connected component). Given two normalized network adjacency matrices $\mathbf{A_1}$ and $\mathbf{A_2}$ characterized by the functions $\beta_{0,1}$ and $\beta_{0,2}$, we then measure their dissimilarity as the absolute area between these two functions (Figure 3):

$$D_{PH}(\mathbf{A_1}, \mathbf{A_2}) = \sum_{w_t=0}^{1} |\beta_{0,1}(w_t) - \beta_{0,2}(w_t)| \tag{2}$$

In order to apply a Louvain community detection algorithm, these dissimilarity measures D are converted into similarity measures using

$$S(\mathbf{A_1}, \mathbf{A_2}) = 1 - \frac{D(\mathbf{A_1}, \mathbf{A_2})}{\max D}. \tag{3}$$

where $\max D$ here gives the maximum dissimilarity observed between any two barcodes from a single year within the entire data set. For each year of data, the "network" of national IO tables (with links weighted

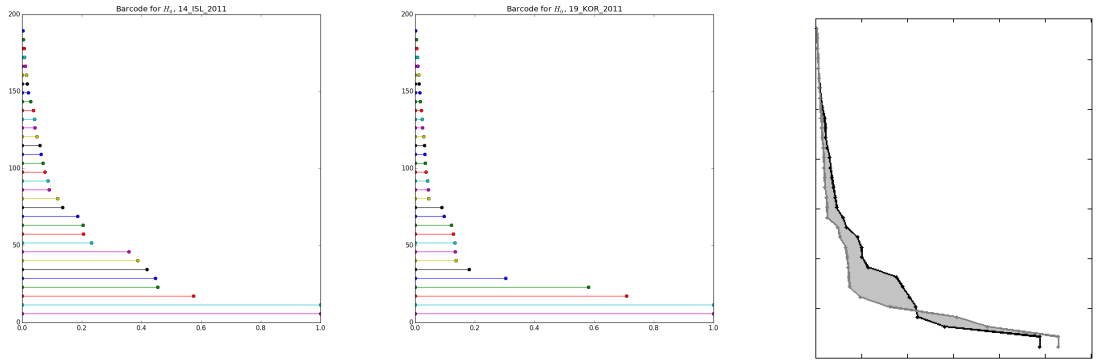

Figure 3: Two β_0 persistent homology barcodes (left and center) are viewed as functions (right), with the total shaded area between the curves serving as a dissimilarity measure.

by the similarity measure) are then treated with a Louvain community detecton algorithm, yielding the clustering with optimal modularity [4].

2.4.2 Network similarity via Pearson correlation

Given two adjacency matrices $\mathbf{A_1}$ and $\mathbf{A_2}$, we compute Pearson correlation as

$$S_{\text{corr}}(\mathbf{A_1}, \mathbf{A_2}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\left(A_{1i,j} - \overline{\mathbf{A_1}}\right)\left(A_{2i,j} - \overline{\mathbf{A_2}}\right)}{\left[\left(A_{1i,j} - \overline{\mathbf{A_1}}\right)^2\left(A_{1i,j} - \overline{\mathbf{A_1}}\right)^2\right]^{\frac{1}{2}}}. \tag{4}$$

This is used directly as a similarity measure in clustering analyses.

2.4.3 Network similarity via Frobenius norm difference

The Frobenius matrix norm can be applied to the difference of two IO table matrices as a distance metric. Considering two IO networks representing N-sector economies, $\mathbf{A_1}$ and $\mathbf{A_2}$, their Frobenius norm distance is

$$D_F(\mathbf{A_1}, \mathbf{A_2}) = \left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left((A_1)_{i,j} - (A_2)_{i,j}\right)^2\right]^{\frac{1}{2}}. \tag{5}$$

This is converted into a similarity measure again using Equation 3.

2.4.4 Survey of cluster means for economic indicators

Data was taken from the database of World Bank Development Indicators [12] for several economic indicators of interest: GDP growth, GDP per capita, natural log of GDP ("Ln(GDP)"), export percentage of GDP ("Export %"), import percentage of GDP ("Import %"), inflation, percent of Internet users in the population ("% Internet users"), life expectancy, and natural log of population ("Ln(Population)"). Mean values of these indicators were computed for each of the clusters resulting from the maximum-modularity stage of a Louvain community detection algorithm applied to each annual data set. For each pair of clusters, p-values gauge

statistical significance of the grouping; low p-values between clusters with regards to a certain indicator suggest that the similarity measure in question tends to dependably distinguish networks in terms of that indicator.

3 Results: Clustering analysis

3.1 Summary of clustering analyses

A Louvain community detection algorithm was applied to each year of OECD domestic Input-Output data using each of the 3 similarity measures discussed above. For the highest-modularity clustering, the average values of various World Bank economic indicators were computed for each cluster (where data was unavailable for certain nations, those nations were omitted from the averages), along with inter-cluster p-values for each pair of clusters. While the total number of clusters present in the highest-modularity clusterings varied from year to year and method to method, in all observed cases, most nations fell into one of the two largest clusters. We thus summarize the results by focusing on the p-value of the largest two clusters for each year, averaging over all available years (1995, 2005, and 2008-2011) for each similarity measure, as shown in Table 1. A more detailed example of results is shown in Appendix B, where the highest-modularity clusterings resulting from each of the three similarity measures, and the corresponding cluster averages for various economic indicators, are displayed for the 2008 OECD IO data set.

Table 1: Mean p-values between two largest clusters and mean modularity values averaged over all years

	Persistent Homology β_0	Pearson correlation	Frobenius distance
GDP growth	.46	.084	.20
GDP per capita	0.60	**.000002******	**.00001******
Ln(GDP)	**.0002*****	.29	.25
Exports %	**0.03***	.11	.10
Imports %	**0.04***	.15	.15
Inflation %	0.40	**.01***	**.01***
% Internet users	0.34	**.002****	**.02***
Life expectancy	0.52	**.00001******	**.00004******
Ln(Population)	**0.007****	.10	.22
Modularity	.071	.080	.050

3.2 Discussion of clustering results

For each annual data set, Louvain clustering using all of the similarity measures tended to produce two primary clusters containing the majority of nations, with several additional clusters containing only a few – sometimes only 1 or 2 – nations. The Persistent Homology similarity measure tended to cleave the set of nations more neatly into fewer groups, consistently yielding 2 large clusters of similar size and several additional small clusters. The highest average modularity value was achieved by the Pearson correlation clusterings (.080), with the Persistent Homology similarity measure performing comparably (.071) and Frobenius difference yielding the lowest average modularity (.050).

The Persistent Homology similarity measure yielded two primary clusters distinguished nations in terms of Ln(GDP) with extreme significance ($p = .0002$), and very significantly in terms of Ln(Population) ($p = .007$). The clusters also differed significantly in terms of export and import percentages of GDP. This similarity measure thus seems to roughly distinguish two primary types of economies via their topological "shapes":

those with larger GDP, larger population, and smaller import/export percentages of GDP, and those with smaller GDP, smaller population, and larger import/export percentages of GDP.

The two "geometric" similarity measures yielded qualitatively similar results. Perhaps predictably, the smaller clusters identified by these measures (which despite being normalized for absolute magnitude, still involve direct sector-by-sector link comparisons) appear to isolate nations which are geographically close or otherwise share specific major industrial sectors in common (Greece / Cyprus, Spain / Portugal, South Korea / Taiwan, Estonia / Latvia / Lithuania, etc.), particular similarities to which the topological perspective was agnostic. Despite their differences, both the Frobenius difference and Pearson correlation similarity measures achieved statistically significant p-values for an identical set of indicators: GDP per capita, Inflation percent, Percent of Internet users, and Life expectancy. The p-values for GDP per capita and Life expectancy were extremely significant ($p < .0001$). These results suggests that these characteristics are more strongly associated with the strength of certain specific sectors within the economy than with certain abstract global connectivity patterns. Interestingly, at least for the indicators used in this study, the sets of indicators successfully distinguished by the topological and "geometric" similarity measures were disjoint.

To understand more precisely just what topological properties the Persistent Homology similarity measure is distinguishing, we take a closer look at the the Persistent Homology clustering for the data from 2008 (Table 2 in Appendix B.1) and examine barcode images for the nations from each cluster. As seen in these representative barcodes (France for Cluster 1 and Bahrain for Cluster 2 as seen in Figure 4), some obvious differences are visible: as visualized previously in Figure 1, France's IO network becomes highly interconnected for lower values of the threshold parameter, as manifest here in its more rapidly-sloping barcode profile, suggesting that the intersectoral flows show a fairly uniformly interconnected structure with link weights distributed in a more "egalitarian" manner throughout the economy. Barcodes of Cluster 2 nations like Bahrain, however, recede more gradually, suggesting the presence of intersectoral flows of more disparate magnitudes and the presence of more distinct clusters of interconnected sectors which are weakly connected to one another. As for other years too, clustering analyses seem to partition the nations in terms of these two primary interconnectivity types; here, the third small cluster seems to contain nations whose barcodes show a rapid initial decay which then tapers more gradually.

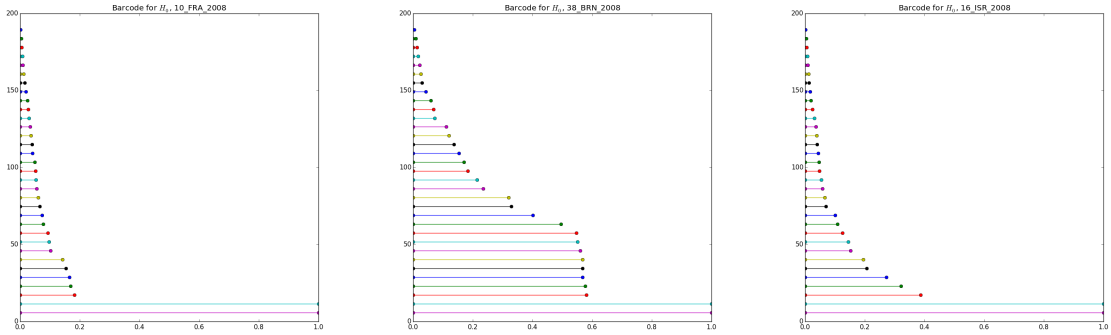

Figure 4: Persistent homology β_0 barcodes for France 2008 (left), Bahrain 2008 (center), and Israel 2008 (right) IO networks

4 Results: Higher-dimensional Persistent Homology barcodes

4.1 Barcodes for dimension 1 and 2 persistent homology

The metrics initially used to compare persistence diagrams involved one-to-one correspondences between barcode features, so that the same number of "holes" must be present in each barcode in order to quantify their similarities. For the zero-dimensional persistent homology, this condition is always trivially satisfied by networks having the number of nodes, since the number of connected components always corresponds to the number of nodes which are unconnected at one extreme of the filtration. However, this is seldom the case for the higher-dimensional PH diagrams studied here. The metric used here to compare the zero-dimensional barcodes, by viewing the barcode profile as the function counting β_0 also does not generalize to higher-dimensional barcodes. Though outside the scope of the current project, some promising possibilities seem to exist for comparing persistence diagrams more generally [5]. In lieu of a similar rigorous clustering analysis for the higher-dimensional barcodes computed, we present some representative higher-dimensional PH barcodes and briefly discuss some of their qualitative features.

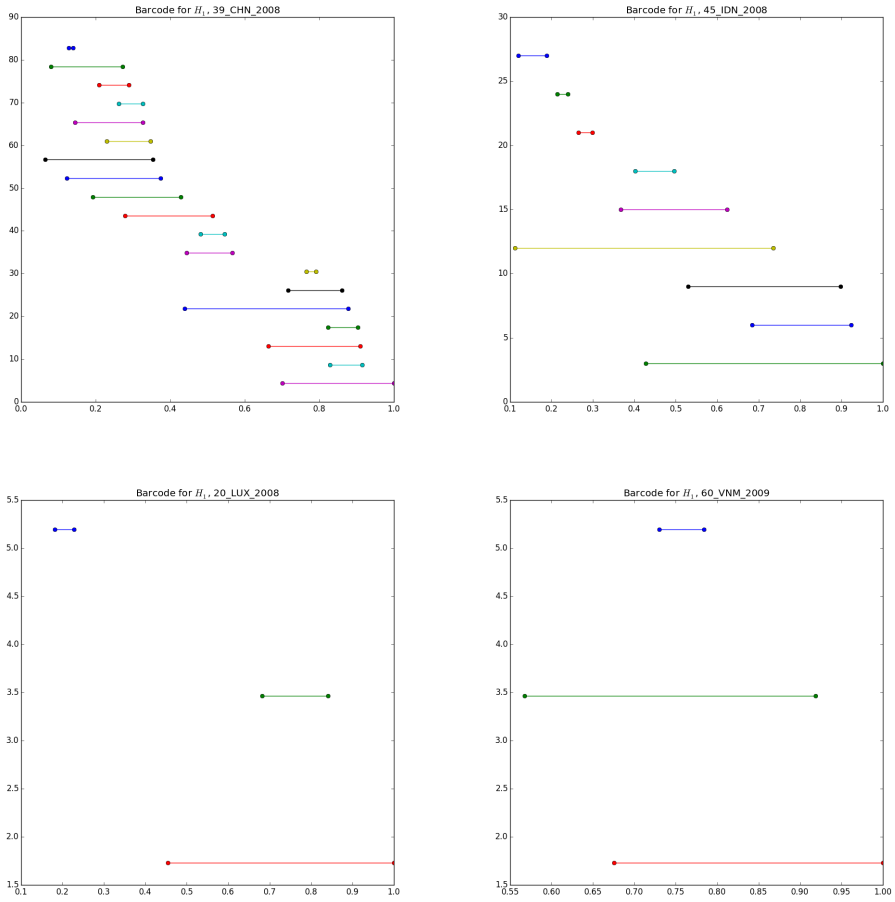

Figure 5: Examples of persistent homology barcodes of dimension 1: Some barcodes feature diagonal cascades of "topological noise" (top left), occasionally interspersed with more persistent holes (top right), while others feature just a few long-lived holes (bottom)

For 1-dimensional "tunnels", barcodes sometimes show a diagonal cascade of so-called "topological noise"

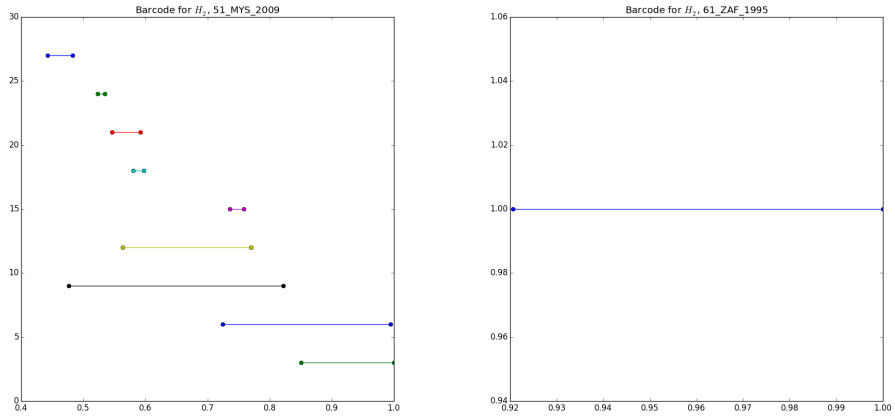

Figure 6: Examples of persistent homology barcodes of dimension 2: While the overwhelming majority of these barcodes are empty, revealing no topological "voids" in their IO networks, several networks reveal a mixture of "topological noise" and persistent voids (left) or a single, shallow void (right).

interspersed by some arguably more persistent features. Others, however, contain just one or two clear, persistent "holes" (see Figure 5).

The majority of dimension 2 barcodes computed are empty, containing no bars corresponding to "voids" (holes of dimension 2, enclosed by 3-cliques of sectors) in the IO network complexes at any point throughout the parameter sweeps. The issue of interpretation of increasingly high-dimensional holes may thus be moot; if the IO data sets studied here are indeed more widely representative of Input-Output data in general, higher-dimensional homological features such as these may be seldom if ever observed in actual IO networks. However, in a few cases, there is some 2-dimensional homological signal observed; for example, a cascade of "topological noise" appears in some cases, and even a "shallow" yet persistent void is observed in others (see Figure 6).

4.2 Exploring the role of "holes" in IO networks

Since Input-Output tables describe the interdependencies of economic sectors, the Input-Output approach has historically been used to study how a disruption or stimulus affecting one economic sector may propagate through other sectors [19]. More recent, explicitly network-theoretical analyses also study the diffusion of "shocks" through IO networks and discuss the implications for economic resilience [11]. It may be natural to begin further explorations into the relevance of the "holes" observed here in IO networks from this perspective: How does the presence of each of these different types of topological holes influence the diffusion of "shocks" through these networks?

5 Conclusion

In this preliminary investigation, we demonstrate that the domestic Input-Output networks representing national economies carry topological and "geometric" hallmarks which signify distinct types of economic properties, even when information about their absolute magnitudes is discarded. Economies with high GDP, large population, and small import/export percentages of GDP exhibit a distinct type of interconnectivity – visible through the Persistent Homology of dimension 0 – from those with lower GDP, small population,

and larger Import/Export percentages, in a way that is invisible to "geometric" perspectives which focus on sector-by-sector comparisons of economic flows. In this way, it offers a complementary perspective to "geometric" similarity measures which more successfully identify economic properties that are associated with the strength of specific sectors within the economies. We also demonstrate that national Input-output networks show rich higher-dimensional persistent-homological "signal" of topological tunnels in dimension 1, and occasionally reveal topological voids of dimension 2, the significance of which may be explored in future work.

Appendix A Introduction to Persistent Homology

A.1 Simplices

A *simplex* is an aggregate of fundamental, irreducible $0-$dimensional objects called *vertices* [13]. Simplices are described by their dimension, or the number of vertices which they associate; an *n-simplex* comprises $(n+1)$ vertices. The sub-simplices, or subsets of $(m+1)$ of these vertices (where $m < n$), are called *faces* of dimension m or *m-faces* of the n-simplex.

So, to use geometric examples, a single vertex x_1 is a $0-$simplex. A $1-$simplex, or a relationship between 2 vertices, can be visualized as a line segment bounded by 2 vertices x_1 and x_2, with the vertices themselves seen as the $0-$faces of the $1-$simplex (x_1x_2). A $2-$simplex $\sigma_2 = (x_1x_2x_3)$ can be visualized as the interior of the triangle defined by three vertices x_1, x_2, and x_3 and thus bounded by the 1-faces (line segments) (x_1x_2), (x_2x_3), and (x_1x_3). A $3-$simplex $(\sigma_3 = x_1x_2x_3x_4)$ can be visualized as the tetrahedron bounded by triangular $2-$faces $(x_1x_2x_3)$, $(x_2x_3x_4)$, $(x_1x_2x_4)$, and $(x_1x_3x_4)$, and so on through arbitrarily high dimensions that are not so straightforward to visualize but follow the same formalism.

We begin by defining these kinds of basic objects because, in general, we can discretize even continuous topological spaces by viewing them as simplicial complexes (for example, studying the topology a 2-D surface by dividing it up into triangular tiles, or studying a 3-D space by segmenting it into tetrahedral compartments). This discretized perspective allows us to deal with the space *combinatorially* – by making reference to the orderings of sequences of these discrete objects and how they are connected with one another – to describe the connectivity and "shape" of the space in a concise and general way. Inherent topological features, such as boundaries or "holes" present in the space, will become visible through the existence of certain types combinations of sequences of these discretized features within the space.

A.2 Simplicial complexes

A *simplicial complex* is a set of simplices of various dimensions and their faces (that is, a *topological space*) for which the following applies:

- All faces of each simplex are also in the set.
- Pairs of simplices intersect only in faces that both simplices share.

Within a simplicial complex, we can consider combinations of simplices that share a common dimension. For example, if σ_i are the k-simplices in the complex, we can construct combinations of these simplices, weighted by integer constants c_i:

$$\sum_i c_i \sigma_i.$$

These are called *k-chains*. The space spanned by these k-chains is denoted as C_k. We also endow these chains with some notion of orientation on the k-simplices, so that if, for example, $\sigma = (x_1 x_2 \dots x_{k-1} x_k)$, then $-\sigma = (x_k x_{k-1} \dots x_2 x_1)$, with the sequences σ composing with its inverse $-\sigma$ as $\sigma + (-\sigma) = 0$.

A.3 Boundary operator

A complex of k-simplices can be specified in terms of all the vertices making up all its simplices, but can also be viewed in terms of the objects which define its boundaries. Vertices, as dimensionless fundamental simplices, have no boundaries. A $1-$simplex $(x_1 x_2)$ is the relationship between those two vertices; in some sense, it consists of the stuff "between" two $0-$dimensional endpoints, and so its boundary can be expressed as the formal difference of these vertices, much as the vector between two positions is expressed as the formal difference of two position vectors: $(x_2 - x_1)$. $2-$simplices (triangles) like $(x_1 x_2 x_3)$ are bounded by the chain giving the sequence of its 1-simplex edges traversed in order: $(x_2 x_3) - (x_3 x_1) + (x_1 x_2)$, and so on.

This correspondence between $(k+1)$-simplices and the k-chains which bound them is given by a mapping called the *boundary operator* $\partial_{k+1} : C_{k+1} \to C_k$. Consistent with the examples given above, its action on a simplex is formally given by

$$\partial_{k+1} \sigma_{k+1} = \partial_{k+1} (x_1 x_2 \dots x_{k+1}) = \sum_j (-1)^{j+1} \widehat{\sigma_j},$$

where $\widehat{\sigma_j} = x_1 \dots x_{j-1} x_{j+1} \dots x_{k+1}$ is the k-simplex made up of all the vertices of σ_{k+1} except x_j. This operator acts on $(k+1)$-chains of simplices linearly. For example, the 1-boundary of a chain of 1-simplices sharing an interior vertex x_2 is given by

$$\partial_1 ((x_1 x_2) + (x_2 x_3)) = \partial_1 ((x_1 x_2)) + \partial_1 ((x_2 x_3)) = [x_2 - x_1] + [x_3 - x_2] = x_3 + (x_2 - x_2) - x_1 = x_3 - x_1,$$

so that the interior vertex is cancelled out, leaving only the outer vertices or "loose ends" defining the boundary of the chain. In general, those k-simplices in the interior will appear twice in the sum with opposite signs so as to cancel out. In this way, the operator acts on simplicial complexes (endowed with some orientation and expressed accordingly as a chain) to pick out the exterior "boundary" simplices of the corresponding dimension and to ignore the shared faces within the interior of the complex.

A.3.1 Boundaries

In abstract algebra terms, the set of all things to which an operator maps objects in its domain is called the *image* of the operator. The boundary operator ∂_{k+1} maps $(k+1)-$chains to their boundaries, which are a subset of the $k-$chains C_k. We denote this set of boundaries as B_k, so

$$B_k \equiv \text{Im } \partial_{k+1}.$$

Focusing on its action on a particular simplicial complex X, we specify

$$B_k(X) \equiv \text{Im } \partial_{k+1}(X).$$

A.3.2 Cycles

If a k-chain, such as the 1-chain above, is connected so that it loops back on itself, with all potential exterior boundary faces now joined together leaving no "loose ends", the chain is called a *cycle* (or, specifying

dimension, a k-cycle). Since these have no leftover "loose ends" to define a k-boundary, k-cycles have no k-boundary, as we can see by applying the boundary operator ∂_1 to a 1-cycle:

$$\begin{aligned}&\partial_1\left((x_1x_2)+(x_2x_3)+(x_3x_1)\right)=\partial_1\left((x_1x_2)\right)+\partial_1\left((x_2x_3)\right)+\partial_1\left((x_3x_1)\right)\\&=[x_2-x_1]+[x_3-x_2]+[x_1-x_3]=(x_1-x_1)+(x_2-x_2)+(x_3-x_3)=0.\end{aligned}$$

In general, the same will hold true for cycles of arbitrary dimension; since all k-simplex terms in the boundary chain of a $(k+1)$-cycle appear twice, with opposite orientation due to the signs specified in the definition of the boundary operator, all k-chains similarly map under the boundary operator to 0, reflecting that the cycle is "closed" and has no boundary.

In abstract algebra terms, the set of all the things that an operator maps to 0 is called the *kernel* of the operator. And so, writing the set of all k-cycles as Z_k, we write

$$Z_k \equiv \text{Ker } \partial_k.$$

A.3.3 Boundaries have no boundaries

The definition of the boundary operator ∂_{k+1} guarantees that all boundaries are in fact k-cycles. These k-cycles have boundary 0 under ∂_k, so that, in general,

$$\partial_k\partial_{k+1}=0.$$

In other words, everything that goes through ∂_{k+1} ends up inside the kernel of ∂_k:

$$\text{Im } \partial_{k+1} \subset \text{Ker } \partial_k \subset C_k,$$

or

$$B_k \subset Z_k \subset C_k.$$

A.4 Homology groups

A.4.1 Quotient groups.

Since the k-boundaries B_k are a subset – actually, a *subgroup* (when we "add" the boundaries from B_k to one another by considering them as a single chain traversed in sequence, the resulting chain still defines a boundary) – of the k-cycles Z_k, they can be used to form a *quotient group* Z_k/B_k. This means that if we consider all the elements in B_k as representing one equivalent element, this grouping yields a natural partition of the larger group Z_k into other elements which respect the original structure (such that for all boundaries $b_k \in B_k$ and other cycles c, then elements cB_k of the new group are formed by the set of all compositions cb_k which satisfy $(c_1B_k)(c_2B_k)=(c_1c_2)B_k$) and so themselves are elements of a more coarsely-defined group; in this way, we clump the details of the group Z_k in some way defined by its zero element B_k, and consider these clumps as the elements of a new group which reflects some some important structure of interest. In the quotient group representing the homology, we shall see that this corresponds to grouping cycles into clumps which enclose the same "holes".

A.4.2 Homology and Betti numbers.

The *n-dimensional homology group* of a simplicial complex X is defined as

$$H_n(X) \equiv \frac{Z_n(X)}{B_n(X)} = \frac{\text{Ker } \partial_n}{\text{Im } \partial_{n+1}}.$$

The *nth Betti number*,

$$\beta_0(X) = \text{rank}\left(H_n(X)\right),$$

counts the number of elements in the corresponding homology group.

A.4.3 0th homology group.

In the case of $H_0 = \frac{\text{Ker } \partial_0}{\text{Im } \partial_1}$, ∂_0 is simply the zero map which sends all vertices to a common element.

The Im ∂_1 is spanned by the boundaries $(v_1 - v_0)$ of all 1-simplices v_0v_1 in the simplicial complex. The quotient group $H_0 = \frac{\text{Ker } \partial_0}{\text{Im } \partial_1}$ thus consists of all vertices in the simplicial complex, but identified with each other by the relations $(v_1 - v_0) = 0$ corresponding to the 1-simplices v_0v_1 present in the complex. In this way, all vertices which can be linked to one another by sequences of adjacent edges are grouped together into a common element in the quotient group, so that the number β_0 of elements in the group corresponds to the number of connected components in the simplicial complex.

A.4.4 Example.

As an example, consider the simplicial complex shown in Figure 7.

For H_0, all chains are cycles, having zero boundary, so that Z_0 is spanned by a,b,c and d. The boundaries B_0 are spanned by the boundaries of the 1-chains: $\partial_1\ (ab) = (b-a)$, $\partial_1\ (bc) = (c-b)$, and $\partial_1\ (ac) = (c-a)$. Thus H_0 consists of all vertices, grouped together by the relations $a = b$, $b = c$, and $a = c$, so that all the vertices form a single zero element. So $\beta_0 = 1$, representing the single connected component of the complex.

For H_1, by inspection it is visible that the cycles Z_1 are spanned by $(ab + bc - ac)$ and $(bc + cd - db)$. The boundaries B_1 are spanned by $\partial_2\ (abc) = (bc - ac + ab)$. The first homology $H_1 = \frac{Z_1}{B_1}$ is thus spanned by the two cycles $(ab + bc - ac)$ and $(bc + cd - db)$, but quotiented by the relation $(bc - ac + ab) = 0$, which identifies the first of these cycles as 0 so that there is a single remaining nonzero element represented by the cycle $(bc + cd - db)$. Thus $\beta_1 = 1$ corresponding to the "tunnel" bounded by the triangular loop on the right-hand side.

A.5 Flag/clique complex

When considered simply as aggregates of 0-simplex vertices linked by 1-simplex edges, graphs would contain at most 0-dimensional homology information. However, there is natural higher-order structure if we associate the presence of an n-vertex *clique* (or *complete subgraph*, in which each of the n vertices in question connected to all of the others) within the graph with an $(n-1)$-simplex in the simplicial complex. The n-way relationship indicated by an n-clique is thus viewed as the higher-order generalization of a 2-way relationship between pairs of vertices indicated by a link. This construction is known as the *flag complex* or *clique complex* of the graph. A "hole" in H_n then indicates a gap bounded by interconnected cliques of $(n + 1)$ vertices.

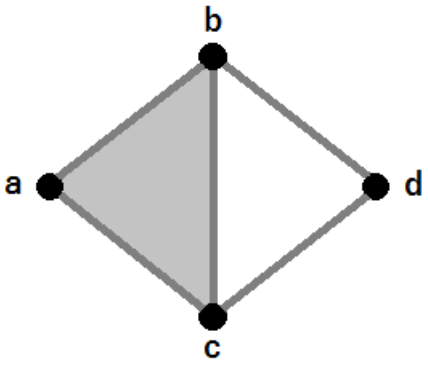

Figure 7: A simplicial complex $\{a, b, c, d, ab, ac, bc, bd, cd, abc\}$

We note that the example shown in Figure 7 is *not* a clique complex, since the 3-clique composed of vertices b, c, and d would be manifest as a 2-simplex bcd, "filling in" the hole bounded by the cycle $(bc + cd - db)$. Thus, a hole of dimension 1 appears in a clique complex only when a cycle of 4 or more 1-simplices bound a hold which is not filled in by a complete clique structure.

Appendix B Selected results from clustering analysis

B.1 Clustering analysis of Persistent Homology β_0 barcodes (2008 data)

Table 2: Clustering partition of Persistent Homology β_0 barcodes for 2008 data (Modularity: 0.073)

Cluster 1	Cluster 2	Cluster 3
2 AUT Austria	1 AUS Australia	9 FIN Finland
3 BEL Belgium	5 CHL Chile	16 ISR Israel
4 CAN Canada	12 GRC Greece	32 TUR Turkey
6 CZE Czech Republic	13 HUN Hungary	58 TUN Tunisia
7 DNK Denmark	14 ISL Iceland	
8 EST Estonia	15 IRL Ireland	
10 FRA France	19 KOR South Korea	
11 DEU Germany	20 LUX Luxembourg	
17 ITA Italy	23 NZL New Zealand	
18 JPN Japan	24 NOR Norway	
21 MEX Mexico	28 SVN Slovenia	
22 NLD Netherlands	35 ARG Argentina	
25 POL Poland	36 BGR Bulgaria	
26 PRT Portugal	38 BRN Bahrain	
27 SVK Slovakia	39 CHN China	
29 ESP Spain	40 COL Colombia	
30 SWE Sweden	41 CRI Costa Rica	
31 CHE Switzerland	42 CYP Cyprus	
33 GBR Great Britain	43 HKG Hong Kong	
34 USA United States	46 IND India	
37 BRA Brazil	47 KHM Cambodia	
44 HRV Croatia	49 LVA Latvia	
45 IDN Indonesia	50 MLT Malta	
48 LTU Lithuania	55 SAU Saudi Arabia	
51 MYS Malaysia	56 SGP Singapore	
52 PHL Philippines	57 THA Thailand	
53 ROU Romania	60 VNM Vietnam	
54 RUS Russia	61 ZAF South Africa	
59 TWN Taiwan		
62 ROW Rest of World		

Table 3: Mean economic indicators by cluster: 2008 data (Persistent Homology β_0)

	Cluster 1	Cluster 2	Cluster 3	Significant p-value
GDP growth	1.8617	2.5984	2.4042	
GDP per capita	24396	21601	19075	
Ln(GDP)	**27.017**	**25.429**	25.837	**0.0007*** **
Exports %	**43.497**	**66.731**	40.946	**0.039***
Imports %	**43.46**	**63.774**	42.062	**0.044***
Inflation %	**5.312**	**8.1509**	6.0071	**0.027***
% Internet users	56.626	47.341	51.24	
Life expectancy	76.695	76.152	77.089	
Ln(Population)	**17.297**	**16.104**	16.377	**0.024***

B.2 Clustering analysis by Pearson correlation (2008 data)

Table 4: Clustering partition by Pearson correlation for 2008 (Modularity 0.084)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
1 AUS Australia	4 CAN Canada	5 CHL Chile	25 POL Poland	42 CYP Cyprus	59 TWN Taiwan
2 AUT Austria	38 BRN Bahrain	14 ISL Iceland		44 HRV Croatia	
3 BEL Belgium	54 RUS Russia	19 KOR South Korea			
6 CZE Czech Republic	61 ZAF South Africa	21 MEX Mexico			
7 DNK Denmark		23 NZL New Zealand			
8 EST Estonia		32 TUR Turkey			
9 FIN Finland		35 ARG Argentina			
10 FRA France		36 BGR Bulgaria			
11 DEU Germany		37 BRA Brazil			
12 GRC Greece		39 CHN China			
13 HUN Hungary		40 COL Colombia			
15 IRL Ireland		41 CRI Costa Rica			
16 ISR Israel		45 IDN Indonesia			
17 ITA Italy		46 IND India			
18 JPN Japan		47 KHM Cambodia			
20 LUX Luxembourg		48 LTU Lithuania			
22 NLD Netherlands		51 MYS Malaysia			
24 NOR Norway		52 PHL Philippines			
26 PRT Portugal		53 ROU Romania			
27 SVK Slovakia		55 SAU Saudi Arabia			
28 SVN Slovenia		57 THA Thailand			
29 ESP Spain		58 TUN Tunisia			
30 SWE Sweden		60 VNM Vietnam			
31 CHE Switzerland		62 ROW Rest of World			
33 GBR Great Britain					
34 USA United States					
43 HKG Hong Kong					
49 LVA Latvia					
50 MLT Malta					
56 SGP Singapore					

Table 5: Mean economic indicators by cluster: 2008 data (Pearson correlation)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Significant p-value
GDP growth	**0.76062**	1.9187	**4.0136**	3.8664	2.8369		**.000009******
GDP per capita	**34876**	18840	**9151.4**	9437.6	18950		**.0000002******
Ln(GDP)	26.442	26.207	26.073	26.609	24.206		
Exports %	65.027	44.925	43.174	38.318	44.327		
Imports %	62.684	29.917	44.308	43.191	53.736		
% Inflation	**4.5209**	7.5249	**9.5075**	4.3494	5.3804		**.00007******
% Internet users	***68.803***	*39.49*	**33.788**	53.13	43.275		**.000000005****** ; *0.0013***
Life expectancy	***79.346***	*69.659*	**73.97**	75.544	77.459		**.0000002****** ; *0.0004****
Ln(Population)	**16.115**	16.665	**17.554**	17.456	14.597		**0.0057****

B.3 Clustering analysis by Frobenius norm difference (2008 data)

Table 6: Clustering partition by Frobenius norm difference for 2008 data (Modularity: 0.049)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
1 AUS Australia	4 CAN Canada	5 CHL Chile	12 GRC Greece	13 HUN Hungary	19 KOR South Korea	25 POL Poland
2 AUT Austria	22 NLD Netherlands	14 ISL Iceland	42 CYP Cyprus		59 TWN Taiwan	53 ROU Romania
3 BEL Belgium	24 NOR Norway	21 MEX Mexico	44 HRV Croatia			54 RUS Russia
6 CZE Czech Republic	34 USA United States	23 NZL New Zealand				62 ROW Rest of World
7 DNK Denmark	38 BRN Bahrain	32 TUR Turkey				
8 EST Estonia	55 SAU Saudi Arabia	35 ARG Argentina				
9 FIN Finland	61 ZAF South Africa	36 BGR Bulgaria				
10 FRA France		37 BRA Brazil				
11 DEU Germany		39 CHN China				
15 IRL Ireland		40 COL Colombia				
16 ISR Israel		41 CRI Costa Rica				
17 ITA Italy		45 IDN Indonesia				
18 JPN Japan		46 IND India				
20 LUX Luxembourg		47 KHM Cambodia				
26 PRT Portugal		48 LTU Lithuania				
27 SVK Slovakia		51 MYS Malaysia				
28 SVN Slovenia		52 PHL Philippines				
29 ESP Spain		57 THA Thailand				
30 SWE Sweden		58 TUN Tunisia				
31 CHE Switzerland		60 VNM Vietnam				
33 GBR Great Britain						
43 HKG Hong Kong						
49 LVA Latvia						
50 MLT Malta						
56 SGP Singapore						

Table 7: Mean economic indicators by cluster: 2008 data (Frobenius norm difference)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Significant p-value
GDP growth	***0.80855***	1.8606	**3.7865**	1.7432	0.87858	2.8292	*4.6139*	**0.0002*** ; *0.0076***
GDP per capita	**34075**	*34595*	***8507.2***	20683	11750	20928	7427.3	**.000002**** ; *0.0009**** ; 0.0044**
Ln(GDP)	26.294	26.865	**25.714**	24.912	25.493	27.655	**27.824**	**0.038***
Exports %	68.684	48.701	43.209	37.34	79.956	49.961	32.388	
Imports %	66.182	34.516	45.307	47.941	79.601	49.972	34.642	
% Inflation	***4.6127***	5.136	**9.8757**	4.9712	6.0662	4.6743	*8.8291*	**0.00038*** ; *0.015***
% Internet users	***68.516***	59.874	**31.911**	41.583	61	81	33.917	**.00000001**** ; *.00001**** ; 0.00071***
Life expectancy	***79.519***	74.887	**73.923**	78.285	73.702	79.833	*71.436*	**.000001**** ; *.000006**** ; 0.040*
Ln(Population)	**15.983**	16.644	**17.351**	15.142	16.122	17.706	18.926	**0.013***

References

[1] Fidel Aroche-Reyes. A qualitative input-output method to find basic economic structures*. *Papers in Regional Science*, 82(4):581–590, 2003.

[2] Asian Development Bank. Urban metabolism of six Asian cities. `http://wcm.adb.org/sites/default/files/publication/59693/urban-metabolism-six-asian-cities.pdf`, 2014. Accessed: 2015-06-29.

[3] Florian Blöchl, Fabian J Theis, Fernando Vega-Redondo, and Eric ON Fisher. Vertex centralities in input-output networks reveal the structure of modern economies. *Physical Review E*, 83(4):046127, 2011.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10): P10008, 2008.

[5] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.

[6] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[7] CJ Carstens and KJ Horadam. Persistent homology of collaboration networks. *Mathematical Problems in Engineering*, 2013, 2013.

[8] V Carvalho. Input-output networks. a survey. *Complexity Research Initiative for Systemic Instabilities. FP7-ICT-2011-7-288501-CRISIS*, 2012.

[9] Vasco M Carvalho. *Aggregate fluctuations and the network structure of intersectoral trade*. ProQuest, 2008.

[10] Federica Cerina, Zhen Zhu, Alessandro Chessa, and Massimo Riccaboni. World input-output network. *PloS one*, 10(7):e0134025, 2015.

[11] Martha G Alatriste Contreras and Giorgio Fagiolo. Propagation of economic shocks in input-output networks: A cross-country analysis. *Physical Review E*, 90(6):062812, 2014.

[12] World Bank Group. World development indicators 2012, 2012.

[13] Allen Hatcher. Algebraic topology. 2002. *Cambridge UP, Cambridge*, 606(9).

[14] César Hidalgo. The dynamics of economic complexity and the product space over a 42 year period. *Center for International Development Working Paper*, (189), 2009.

[15] Danijela Horak, Slobodan Maletić, and Milan Rajković. Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03034, 2009.

[16] Hyekyoung Lee, Hyejin Kang, Moo K Chung, Bung-Nyun Kim, and Dong Soo Lee. Persistent brain network homology from the perspective of dendrogram. *Medical Imaging, IEEE Transactions on*, 31 (12):2267–2277, 2012.

[17] Wassily W Leontief. *Input-output economics*. Oxford University Press on Demand, 1986.

[18] James McNerney, Brian D Fath, and Gerald Silverberg. Network structure of inter-industry flows. *Physica A: Statistical Mechanics and its Applications*, 392(24):6427–6441, 2013.

[19] Ronald E Miller and Peter D Blair. *Input-output analysis: foundations and extensions*. Cambridge University Press, 2009.

[20] Organisation for Economic Co-operation and Development. OECD Inter-Country Input-Output (ICIO) Tables, edition 2015: Access to data. http://www.oecd.org/sti/ind/input-outputtablesedition2015accesstodata.htm, 2015. Accessed: 2015-06-29.

[21] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *arXiv preprint arXiv:1506.08903*, 2015.

[22] Sony Pellissery. Viewing economy as a network: An exploration through input-output model. *Available at SSRN 1011363*, 2007.

[23] G Petri, P Expert, F Turkheimer, R Carhart-Harris, D Nutt, PJ Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101): 20140873, 2014.

[24] Giovanni Petri, Martina Scolamiero, Irene Donato, and Francesco Vaccarino. Topological strata of weighted complex networks. *PloS one*, 8(6):e66506, 2013.

[25] P Slater. The network structure of the United States input-output table. *Empirical Economics*, 3(1): 49–70, 1978.

[26] Norihiko Yamano and Nadim Ahmad. The OECD input-output database. 2006.

[27] Yan Zhang, Zhifeng Yang, and Xiangyi Yu. Ecological network and emergy analysis of urban metabolic systems: model development, and a case study of four Chinese cities. *Ecological Modelling*, 220(11): 1431–1442, 2009.

[28] Xiaoyue Zhao. *Sector Similarity in Input-Output Networks*. PhD thesis, University of Michigan, 2015.