

PAPER • OPEN ACCESS

## The different structure of economic ecosystems at the scales of companies and countries

To cite this article: Dario Laudati *et al* 2023 *J. Phys. Complex.* 4 025011

View the [article online](#) for updates and enhancements.

You may also like

- [Measuring the pitch control of professional football players using spatiotemporal tracking data](#)

Lewis Higgins, Tobias Galla, Brian Prestidge et al.

- [Analytic solution of the resolvent equations for heterogeneous random graphs: spectral and localization properties](#)

Jeferson D Silva and Fernando L Metz

- [Hyper-diffusion on multiplex networks](#)

Reza Ghorbanchian, Vito Latora and Ginestra Bianconi



## PAPER

## OPEN ACCESS



RECEIVED  
1 July 2022ACCEPTED FOR PUBLICATION  
6 April 2023PUBLISHED  
18 May 2023

Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



# The different structure of economic ecosystems at the scales of companies and countries

Dario Laudati<sup>1</sup>, Manuel S Mariani<sup>2,3</sup> , Luciano Pietronero<sup>4</sup> and Andrea Zaccaria<sup>5,\*</sup> <sup>1</sup> Dipartimento di Fisica, Sapienza Università di Roma, 00185 Rome, Italy<sup>2</sup> Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China<sup>3</sup> URPP Social Networks, Universität Zürich, 8050 Zürich, Switzerland<sup>4</sup> Centro Ricerche Enrico Fermi, 00184 Rome, Italy<sup>5</sup> Istituto dei Sistemi Complessi, UOS Sapienza, CNR, 00185 Rome, Italy

\* Author to whom any correspondence should be addressed.

E-mail: [andrea.zaccaria@cnr.it](mailto:andrea.zaccaria@cnr.it)**Keywords:** economic complexity, companies, nestednessSupplementary material for this article is available [online](#)

## Abstract

A key element to understand complex systems is the relationship between the spatial scale of investigation and the structure of the interrelation among its elements. When it comes to economic systems, it is now well-known that the country-product bipartite network exhibits a nested structure, which is the foundation of different algorithms that have been used to scientifically investigate countries' development and forecast national economic growth. Changing the subject from countries to companies, a significantly different scenario emerges. Through the analysis of a unique dataset of Italian firms' exports and a worldwide dataset comprising countries' exports, here we find that, while a globally nested structure is observed at the country level, a local, in-block nested structure emerges at the level of firms. This in-block nestedness is statistically significant with respect to suitable null models and the algorithmic partitions of products into blocks correspond well with the UN-COMTRADE product classification. These findings lay a solid foundation for developing a scientific approach based on the physics of complex systems to the analysis of companies, which has been lacking until now.

Understanding the structure of interactions in a complex system is a fundamental issue [1, 2], since the structure affects the system's function [3, 4] and its resilience against diverse perturbations [5–8]. Yet, interactions can be bounded by different kinds of constraints [9]. When this is the case, understanding the structure and dynamics of interactions requires to identify clear boundaries that separate an ecosystem from its surroundings. While this idea and the resulting methods [9–12] have found promising initial applications in ecological [13–16], neural [17, 18] and social networks [19, 20], they have not yet been applied to economic systems where actors produce and export products. As for these systems, most studies assume that the ecosystem where a country operates is the entire world [21–23]: in principle, each country competes with all the others, and all products are considered. To uncover the complexity of countries' export structure, the world trade web is often represented as a bipartite network where countries and products constitute the nodes of the two layers [24, 25]. At this global scale, a peculiar property emerges: nestedness [12]. Well-known in ecology, in this context nestedness means that developed countries are highly diversified and produce all kinds of products, while developing countries only produce a few ubiquitous products. This empirical observation led to the development of economic complexity, an interdisciplinary approach which applies methods from statistical physics and network science to uncover the determinants of country development [24, 25]. Notably, a predictive approach based on nestedness [25] is able to forecast GDP growth with a significant improvement and complementarity with respect to the IMF projections [26, 27].

Despite these remarkable achievements, a fundamental question remains still open: given that the export of countries is nothing more than the result of the production of individual companies at national level, can the economic complexity approach be extended to the scale of companies? In principle, answering this question could provide both firms and policymakers with an algorithmic tool to evaluate the competitiveness of a company, to design optimal strategies for development, and to forecast its economic performance. Indeed, a high-fitness firm will likely be able to exploit countries' comparative advantage.

Despite its importance, this investigation has been hindered by two main factors. First, data scarcity: export data at the company level is extremely sensitive in terms of privacy policy and is much less homogeneous with respect to the harmonized data about the international trade. Second, and more importantly, the networks of countries and companies may have different structures and, regarding firms, it is unclear how to detect a suitable economic environment—or ecosystem—wherein to assess a firm's competitiveness. Note that such ecosystem could be composed of possible competitors but also suppliers or service providers, as long as a set of capabilities are shared, in the spirit of Porter's theory of management [28, 29].

Thanks to our collaboration with the Italian National Institute of Statistics (IT ISTAT), we could overcome these limitations and access a unique dataset of Italian firms' export records. The products are coded in the same way as previously-analysed datasets of countries' exports, which enables a direct comparison of the structures of the country-product and company-product ecosystems.

Building on this dataset, we apply algorithms to statistically validate the presence of modularity, nestedness, and in-block nestedness [11, 20] to both the country-product and the company-product networks. We find that the same level of nestedness which is present at the country scale is absent when one looks at a national economy of companies as a whole, but re-emerges at the local level, once that the modular structure of the company-product network is considered. As a result of these structural differences, the ranking algorithms developed to evaluate countries' competitiveness do not work properly when applied to the network of companies as a whole. At the same time, the detected in-block nestedness of the company-product network opens up the possibility to apply the economic complexity framework also at the company level, provided that the proper locally-nested ecosystems are considered: the company, its competitors, and the products they compete on.

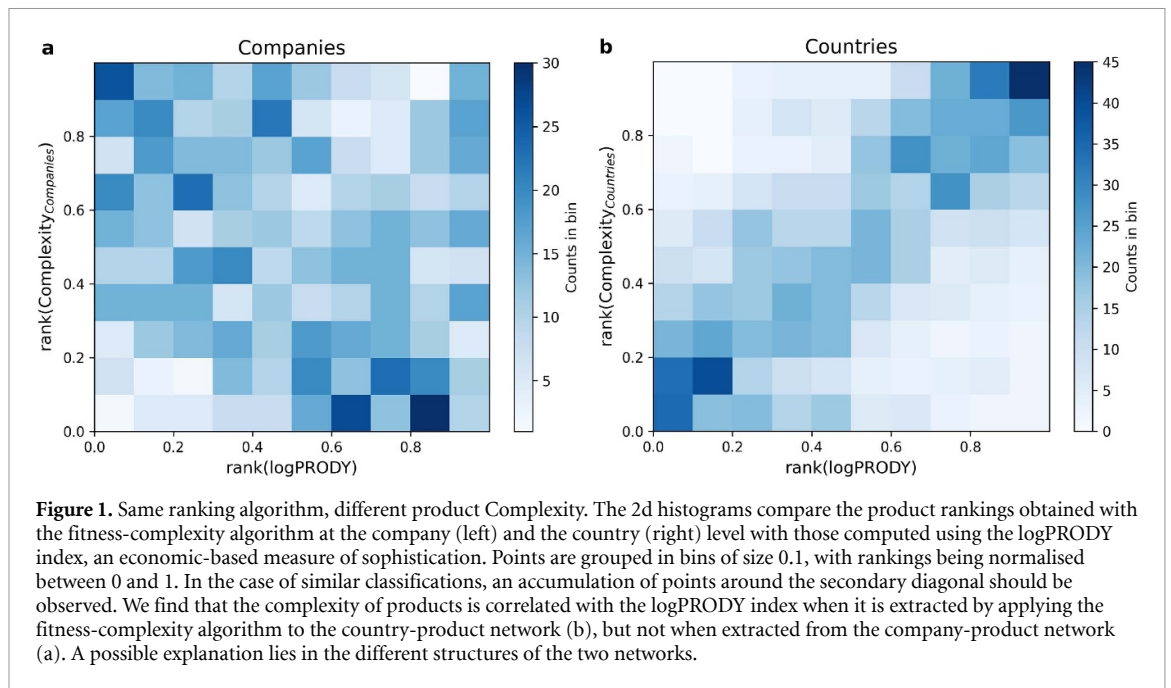
## 1. Results

### 1.1. Same ranking algorithm, different conclusions

Economic complexity algorithms were originally designed to evaluate the competitiveness of countries and the complexity of products from the structure of the country-product network [24, 25]. We begin by showing that state-of-the-art economic complexity algorithms are inadequate to capture the complexity of products and the competitiveness of firms from the structure of the company-product network [30]. To demonstrate this point, we apply the fitness-complexity (FC) algorithm [25, 31] (see Methods for the mathematical formulation) to both the country-product and the company-product networks. In this way, we obtain two different evaluations of the same quantity: the complexity of products. To assess whether the obtained complexity scores are good proxies for the economic value of a product, we compare the complexity rankings with those obtained according to the logPRODY index (see Methods), an external monetary metric that measures the sophistication of products from the GDP of the exporting countries [32, 33]. We expect that a reasonable measure of complexity should exhibit a good correlation with logPRODY.

A good agreement between complexity and logPRODY is only observed when the complexity score is obtained by applying the FC algorithm to the country-product network (Spearman's correlation coefficient  $\rho = 0.642$ ), but not when the same algorithm is applied to the company-product network ( $\rho = -0.224$ , see figure 1). A similar conclusion is reached by comparing the countries' and companies' (extensive) fitness scores with their export volumes, which can be interpreted as proxies for their competitiveness. We only observe a high correlation for countries ( $\rho = 0.887$ ), but not for companies ( $\rho = 0.378$ ) - see supplementary section II and [30] for a comparison with the degree. These results indicate that when applied to the company-product network, the FC algorithm does not accurately estimate the economic value of products and the competitiveness of the exporters. It comes therefore natural to wonder why the FC algorithm can not be applied as it is on the company-product network; our starting point is the crucial requirement that the whole structure of the network should be nested.

Indeed, a possible answer lies in the different structures of the company-product and country-product networks. The FC algorithm builds on the premise that competitive countries tend to diversify their export baskets as much as possible, given their available capabilities [34–36]. This hypothesis is motivated by the globally nested structure of the country-product network [12, 25]: the most diversified countries export all kinds of products, whereas the products exported by more specialised countries are typically exported by the



diversified ones as well. Drawing a parallel with ecology, a more diversified export basket might increase the robustness of a country's economy with respect to adverse external events [6, 37]. A similar argument might, in principle, apply to firms as well, and prior works have associated the diversification of a firm's activities with its economic performance [38–40]. Yet, other studies emphasised the importance for a firm to diversify within its core set of capabilities, avoiding unrelated activities [30, 41–45]. These works suggest that a globally nested structure might not be found, and that the company-product network might be instead partitioned into specialised blocks, which would disagree with the basic premise of the FC algorithm.

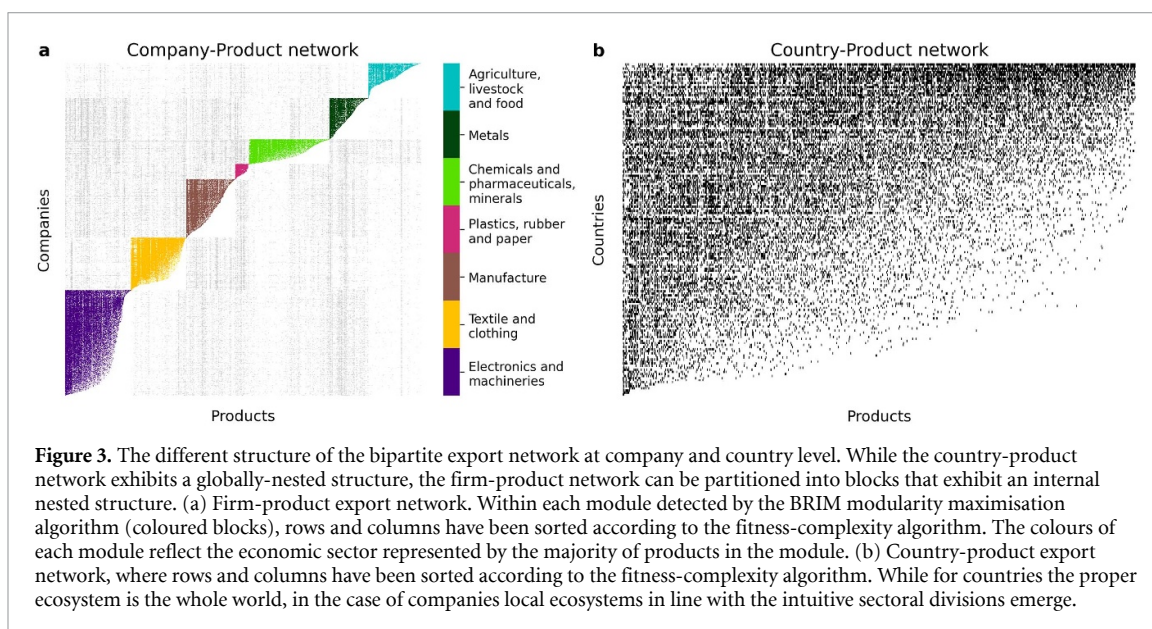
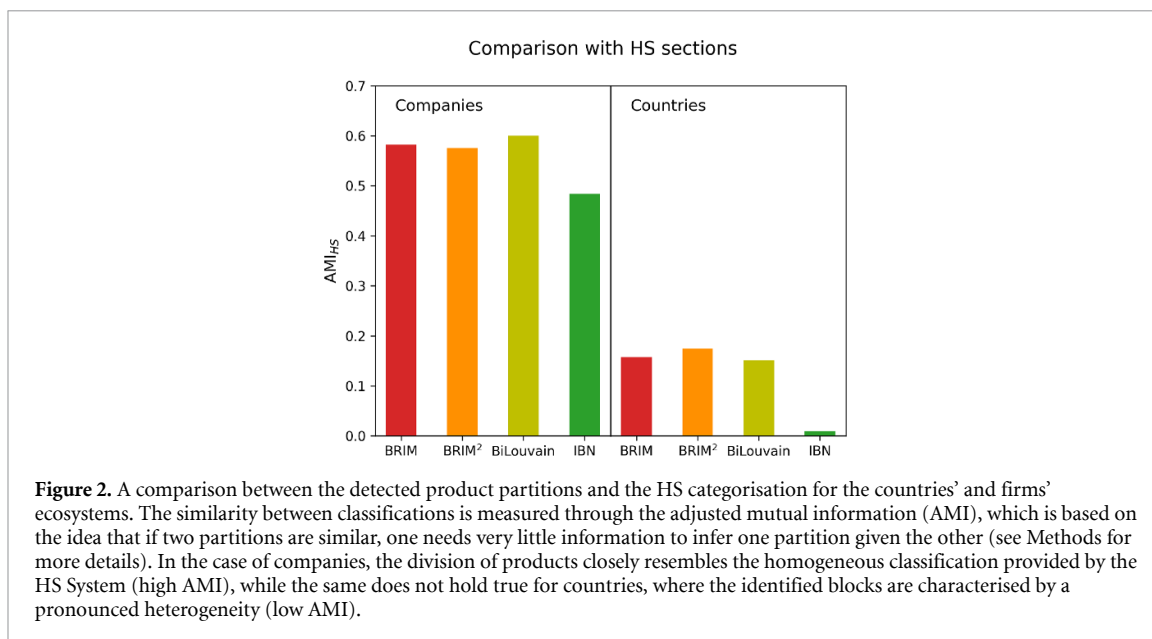
## 1.2. The different role of modularity

The ultimate test of these conjectures lies in the empirical data. To identify potential differences in the structure of the two economic systems, we apply a modularity maximisation algorithm (BRIM, see Methods for more details) to both. By only looking at the modularity scores of the two networks, one may naively conclude that both exhibit a pronounced modular structure ( $Q = 0.218$ ,  $p < 0.01$  for the country-product network,  $Q = 0.512$ ,  $p < 0.01$  for the company-product network, where the  $p$ -values have been obtained with the BiCM null model, see Methods).

Yet there are two substantial differences between the two modular patterns. First, the detected partitions are much noisier in the country-product network than in the company-product network (see Supplementary figure S2 for a visual comparison). More specifically, the blocks in the country-product network contain only 50% of the links, whereas in the company-product network they contain more than 70% of the links.

Second, the interpretation of the detected blocks is radically different in the two systems. To interpret the detected blocks, we investigate their sector composition. To this end, we compare the modularity-detected partition of products with the ones corresponding to the 21 sections of the official export classification, that is the Harmonized System (HS), 1992 edition (see supplementary section V for a detailed description of this classification). The basic idea is that, since the HS sections represent homogeneous categories of products, then coherent (specialised) partitions should show a substantial degree of relatedness, that is, similar products should co-occur in the same blocks. Conversely, heterogeneous (diversified) partitions should show a low degree of relatedness. To ensure the robustness of our conclusions, we perform the comparison between HS sections and the partitions extracted by several different community detection algorithms (BRIM and BRIM<sup>2</sup> [46], BiLouvain [47], and IBN [11]—see Methods for more details): robust results should not depend on the particular algorithm employed, as long as it provides a reasonable partition. The similarity between the partitions is measured using the adjusted mutual information [48] (AMI, see Methods for the mathematical definition). We emphasise that this analysis does not aim to evaluate the detected partitions [49], but only to provide a robust interpretation of the detected modules.

We find that the AMI is significantly larger in the firm-product than in the country-product network. For example, by using the Bipartite, Recursively Induced Modules (BRIM) algorithm, the AMI is 270% larger in the company-product network than in the country-product network. Qualitatively similar results hold for

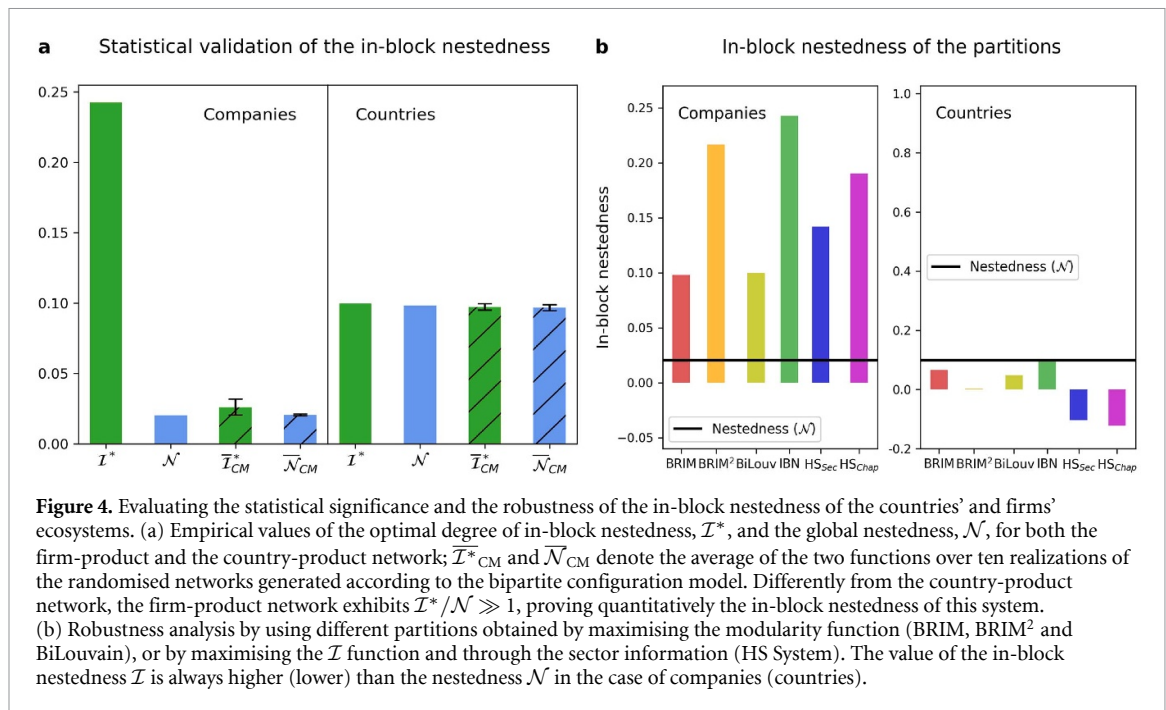


other algorithms (see figure 2). This set of results indicates that companies are genuinely specialised entities that mostly focus on homogeneous groups of products. By contrast, countries do not specialise in confined groups of similar products, as proven by the high heterogeneity of the detected blocks: developed countries diversify their production [25]. It can be shown that the significant degree of modularity observed in the country-product network can be explained by the countries' diversification patterns (see supplementary section III B for a detailed discussion).

These results lead to the investigation of the internal structure of the detected company-product blocks. In particular, the interesting point to evaluate is whether there is a resemblance between the structure of these blocks and the global structure that characterises the country-product network. Such evidence would support the idea that the detected blocks act as boundaries that limit the companies' ability to diversify. To this end, we apply the FC algorithm to the BRIM blocks, and we use the rankings to order the company-product matrices. The result is depicted in figure 3(a). Besides a good agreement with the industrial sectors, the blocks identified in the company-product network display another very interesting feature: they exhibit an internally nested structure. This property will be deeply investigated in the next section.

In light of these results, an initial characterisation of the two economic ecosystems can be outlined. For countries, nestedness is the dominant property, whereas modularity (although statistically significant) emerges only as a second-order feature, essentially determined by the countries' diversification. For companies, modularity is the dominant property, whereas nestedness is relegated within blocks, as a local





property. We can then argue that the reason why the FC algorithm, if applied to the whole company-product network, misestimates the sophistication of products and the competitiveness of companies is that it neglects the block structure of the network.

### 1.3. Local nestedness in the firms' ecosystem

A full validation of the previous characterisation of the two systems requires the deployment of methods that can disentangle the role of nestedness and modularity [11]. Specifically, to prove the claim that nestedness is a local (global) property in the firms' (countries') ecosystems, we implement a recent method to rigorously determine whether a network can be partitioned into blocks with an internal nested structure [11]. This method relies on a quality function—referred to as *in-block nestedness*,  $\mathcal{I}$  (see Methods for the mathematical definition) – and requires to optimise the in-block nestedness function and to compare its optimal value,  $\mathcal{I}^*$ , against the value of the same function for a single-block partition, which we refer to as  $\mathcal{N}$  (see Methods). Large values of the ratio  $\mathcal{I}^*/\mathcal{N}$  indicate that nestedness is a local property, while networks where nestedness is a global property exhibit  $\mathcal{I}^* \simeq \mathcal{N}$  [11, 20].

Our findings quantitatively confirm the qualitative representation of figure 3. We find a large  $\mathcal{I}^*/\mathcal{N}$  ratio for firms ( $\mathcal{I}^*/\mathcal{N} \simeq 12.0$ ; see figure 4(a)), where the in-block nestedness maximisation produces a partition with more than 80 blocks, but not for countries ( $\mathcal{I}^*/\mathcal{N} \simeq 1.02$ ), where only two modules are detected, of which the largest one includes the vast majority of the network nodes (97.6%) - see supplementary figure S5 for a visual representation. To rule out the possibility that large  $\mathcal{I}^*$  values arise through random fluctuations [50], we compare the observed values of  $\mathcal{I}^*$  against those obtained in randomised bipartite networks that preserve on average the nodes' degree (see Methods). We find that the in-block nestedness  $\mathcal{I}^*$  of the firm-product network is significantly larger than that of the corresponding randomised network, whereas the same does not hold for the country-product network (figure 4(a)), where the level of in-block nestedness is entirely due to the degree of global nestedness. Note that by testing the significance of this result with the BICM model, we are performing a highly conservative statistical validation, which can notoriously rule out global nestedness in most empirical networks [51, 52]. Taken together, these results demonstrate that nestedness is a global property for countries, while it emerges locally for firms.

This conclusion is robust with respect to alternative partitions of the network. Specifically, the empirical result that  $\mathcal{I} \gg \mathcal{N}$  holds not only for the optimal in-block nested partition ( $\mathcal{I} = \mathcal{I}^*$ ), but also for reasonable alternative partitions determined by modularity maximisation (via the BRIM, BRIM<sup>2</sup> and BiLouvain methods) or economic sectors (based on HS sections, HS<sub>Sec</sub>, and chapters, HS<sub>Chap</sub>) – see Methods for a summary of these partitioning methods. Although the value of  $\mathcal{I}$  for these partitions is smaller than  $\mathcal{I}^*$ , it remains considerably larger than  $\mathcal{N}$  (see figure 4(b)) – the fingerprint of a network where nestedness is a local network property, and not a global one. Remarkably, in the country-product network, none of the sub-optimal partitions achieves a value of  $\mathcal{I}$  comparable to  $\mathcal{N}$ : we observe  $\mathcal{I} \ll \mathcal{N}$  for all partitions but the

optimal one (for which  $\mathcal{I} = \mathcal{I}^* \simeq \mathcal{N}$ ; see figure 4(b)). This further confirms that nestedness is a global property of the country-product network.

## 2. Discussion

Despite recent advances in economic complexity, comparing the structure and dynamics of economic ecosystems at the country and company scales remained elusive, mostly due to the scarcity of datasets on firms' export activities and the lack of specific methodologies. Here, we overcame this limitation by analysing a unique dataset of Italian firms' exports and a worldwide dataset of the export flows between countries, and by comparing the observed structure of the firms' and countries' ecosystems via recently introduced approaches [11, 20].

Our results reveal that, when looking at an economic ecosystem at different scales, stark structural differences emerge - which is not totally unexpected, given the different objectives of a firm's management with respect to the public sector. More in detail, while we observed a globally nested structure at the country level, we found an in-block nested structure at firm level. We showed that the observed structural differences have profound implications for economic complexity rankings: the FC algorithm [25], which provides an optimal ranking for nested networks, neglects the block structure of interactions and, as a result, it correctly extracts the economic value of products and the competitiveness of economic agents in the country-product network, but not in the firm-product network.

Nevertheless, developing economic-complexity ranking and recommendation algorithms tailored to firms would have profound implications for managerial and policy-making decisions, innovation strategies and investments. To this end, our findings suggest that the first crucial step should be the identification of the local ecosystem of the firms of interest and its boundaries. The appropriate context is not the entire network (as for countries), but is provided by the company-product blocks where the firms operate. Interestingly, since these local ecosystems are internally nested, then applying locally the FC algorithm may still be an effective strategy to rank companies and products within their ecosystem. This analysis will be the subject of future works.

## 3. Methods

### 3.1. Data and network construction

We analysed two datasets: (1) the country-level dataset obtained from the UN-Comtrade dataset (<https://comtrade.un.org>), which is the standard database used in the Economic Complexity framework and (2) the ISTAT dataset concerning the export of Italian companies. In both datasets the export flows are recorded, and products are classified according to a six digit code which, after a data cleaning procedure, was standardised to the HS 1992 categorisation. We then coarse-grained the obtained classification by considering only the first 4 digits, resulting in a set of about 1200 products.

The firms' dataset spans from 1993 to 2017 and it includes 879 280 companies. From year to year the number of companies exporting at least one product varies between 150 000 and 200 000. The countries' dataset spans from 1996 to 2018, and it includes 169 countries in total.

To perform a coherent analysis for both firms and countries we summed up the export volumes for all the available years and only kept the firms (countries) that remained active (for which data is available) for the entire time interval considered. As a result of this filtering procedure, a total of 18 349 firms and 161 countries were left. From these filtered data, we constructed the country-product and the firm-product bipartite binary networks.

The criterion adopted in order to decide whether a country (company) can be considered or not as a competitive exporter of a particular product is the so-called revealed comparative advantage (RCA) [53]. For a pair  $(i, \alpha)$  composed of a potential exporter  $i$  (country or company) and a product  $\alpha$ , the RCA is defined in terms of the ratio between the fraction of export of product  $\alpha$  by country (company)  $i$  and the overall export of  $\alpha$ . The obtained quantity is then divided by the ratio between the total export of  $i$  and the overall export by all countries (companies). This is the most natural way to remove trivial dependencies from the sizes of the economic agents and sectors. In formulas:

$$RCA_{i\alpha} = \frac{\frac{q_{i\alpha}}{\sum_{i'} q_{i'\alpha}}}{\frac{\sum_{\alpha'} q_{i\alpha'}}{\sum_{i'\alpha'} q_{i'\alpha'}}}. \quad (1)$$

As in previous works [24, 25], a threshold value  $R^* = 1$  is used. As a result, a binary country (company)-product matrix  $\mathbf{M}$  is built, whose generic element is:

$$M_{i\alpha} = \begin{cases} 1 & \text{if } RCA_{i\alpha} \geq R^* = 1 \\ 0 & \text{if } RCA_{i\alpha} < R^* = 1 \end{cases}, \quad (2)$$

i.e. country (company)  $i$  can be considered a competitive exporter of product  $\alpha$  if and only if  $M_{i\alpha} = 1$ . In the equivalent network representation, the node of the country (company)  $i$  is linked to the node of the product  $\alpha$  if and only if  $M_{i\alpha} = 1$ . For the characterisation of the basic properties of the two constructed networks, see Supplementary table S1.

### 3.2. Network analysis methods

#### 3.2.1. Modularity

We search for a (sub)optimal modular partition of the nodes by applying a variant of the BRIM algorithm<sup>6</sup> [46] to maximise Barber's modularity [54], defined as:

$$Q = \frac{1}{E} \sum_{i=1}^{N_R} \sum_{\alpha=1}^{N_C} (M_{i\alpha} - P_{i\alpha}) \delta(a_i, a_\alpha), \quad (3)$$

where  $E$  is the number of interactions (links) in the network,  $M_{i\alpha}$  is the biadjacency matrix which denotes the existence of a link between row nodes  $i$  and column nodes  $\alpha$ ,  $P_{i\alpha} = k_i k_\alpha / E$  is the probability that a link between nodes  $i$  and  $\alpha$  exists by chance under a degree-preserving null model,  $a_i$  is a membership variable that defines the block to whom the node  $i$  belongs, and  $\delta(a_i, a_\alpha)$  is the Kronecker delta function, which takes the value 1 if nodes  $i$  and  $\alpha$  are in the same community, and 0 otherwise.

Given the resolution limit that affects modularity optimisation [55], we also considered an alternative method that applies the BRIM algorithm twice, by performing community detection within the blocks identified through the first application of the algorithm. We refer to this method as BRIM<sup>2</sup>. In order to verify the robustness of the results, all the analyses were replicated using the BiLouvain algorithm, which is the extension to bipartite networks of the popular Louvain algorithm introduced by Blondel *et al* [47].

#### 3.2.2. Global and In-block nestedness

In-block nested structures are patterns of interactions characterised by compartments of nodes that internally exhibit a nested pattern of interactions. Using the formulation developed in [11], the degree of in-block nestedness  $\mathcal{I}$  of a network can be quantified as:

$$\mathcal{I} = \frac{2}{N_R + N_C} \left\{ \sum_{i,j} \frac{O_{ij} - \langle O_{i,j} \rangle}{k_j (C_i - 1)} \Theta(k_i - k_j) \delta(a_i, a_j) + \sum_{\alpha,\beta} \frac{O_{\alpha,\beta} - \langle O_{\alpha,\beta} \rangle}{k_\beta (C_\beta - 1)} \Theta(k_\alpha - k_\beta) \delta(a_\alpha, a_\beta) \right\}, \quad (4)$$

where  $C_i$  is the number of nodes that belong to the block to whom the node  $i$  belongs,  $O_{i,j}$  measures the degree of links overlap between rows node pairs,  $\langle O_{i,j} \rangle$  represents the expected number of links between row nodes  $i$  and  $j$  in the null model and is equal to  $\langle O_{i,j} \rangle = k_i k_j / N_R$ , and  $\Theta(\cdot)$  is the Heaviside step function that guarantees that the overlap is computed only between pair of nodes such that  $k_i > k_j$ . The function  $\mathcal{I}$ , called in-block nestedness fitness, can be interpreted as a generalisation of the *global nestedness* function:

$$\mathcal{N} = \frac{2}{N_R + N_C} \left\{ \sum_{i,j} \frac{O_{ij} - \langle O_{i,j} \rangle}{k_j (N_R - 1)} \Theta(k_i - k_j) + \sum_{\alpha,\beta} \frac{O_{\alpha,\beta} - \langle O_{\alpha,\beta} \rangle}{k_\beta (N_C - 1)} \Theta(k_\alpha - k_\beta) \right\}, \quad (5)$$

introduced in [11] as an overlap-based metric, inspired by the nestedness metric based on overlap and decreasing fill [56], which compares the observed level of nestedness with the expected value under a suitable null model. Noteworthy, the objective function  $\mathcal{I}$  reduces to  $\mathcal{N}$  if one considers a single block

<sup>6</sup> <https://github.com/genisott/pycondor>.



( $a_i = a_\alpha = a, \forall i, \alpha$ ). Here we search for a (sub)optimal in-block nested partition of the nodes by applying a variant of the extremal optimisation algorithm [57], adapted to maximise the in-block nestedness function<sup>7</sup>.

### 3.2.3. Null models and statistical tests

To statistically validate the degree of modularity and in-block nestedness, we have used the bipartite configuration model (BiCM) [23, 58] paired with the  $p$ -value.

The BiCM is an entropy-based and unbiased null model which preserves, on average, the degree of both rows and columns<sup>8</sup>.

The  $p$ -value is computed by measuring the frequency of matrices in the null ensemble that are more modular/in-block nested than the input matrix and a threshold value  $\lambda = 0.05$  is used to denote a statistically significant level ( $p < \lambda$ ). For matrices where no randomised networks satisfy this condition, we conservatively assigned  $p < 1/R$ , where  $R$  is the number of independently generated random matrices.

### 3.2.4. Sectoral partitions

In addition to community detection methods based on maximising modularity and in-block nestedness, we also constructed partitions following the HS classification for products (supplementary section V). In particular, in one case (referred to as HS<sub>Sec</sub>) we partitioned the products according to the 21 HS sections and then we assigned countries (companies) to the block corresponding to their highest export volume. The second method (referred to as HS<sub>Chap</sub>) follows the same strategy, except that the product communities do not correspond to the 21 HS sections but to the 99 HS chapters.

### 3.2.5. Partition similarity measures

To evaluate and compare the performances of the clustering algorithms, here we make use of similarity measures based on information theory, which are built on the idea that if two partitions are similar, one needs very little information to infer one partition given the other, and thus this extra information can be used as a measure of dissimilarity. In particular, we employ the so-called AMI [48], defined as:

$$AMI = \frac{I(X, Y) - E\{I(X, Y)\}}{\frac{1}{2}[H(X) + H(Y)] - E\{I(X, Y)\}}, \quad (6)$$

where  $X$  and  $Y$  are two clusterings,  $I(X, Y)$  is their mutual information and  $E\{I(X, Y)\}$  is a correction for randomness using the permutation model [59], in which clusterings are generated randomly subject to having a fixed number of clusters and points in each cluster. Specifically, the AMI equals 1 when the two clusterings are identical, and 0 when the mutual information between the two clusterings equals its expected value.

## 3.3. Economic complexity methods

### 3.3.1. The FC method

The FC method is a non-linear, iterative approach for economic complexity evaluation [25]. Grounded on the nested network structure of the country-product network, the fitness of a country  $F_i$  is measured by the sum of its exported products, weighted by their complexity  $Q_\alpha$ , while the complexity of a product is measured in a nonlinear way. The underlying intuition is that the information that a product is made in some scarcely competitive countries is sufficient to conclude that the complexity of such product is low. In formulas<sup>9</sup>[25]:

$$\begin{cases} \tilde{F}_i^{(n)} = \sum_\alpha M_{i\alpha} Q_\alpha^{(n-1)} \\ \tilde{Q}_\alpha^{(n)} = \frac{1}{\sum_i M_{i\alpha} \frac{1}{\tilde{F}_i^{(n-1)}}} \end{cases} \longrightarrow \begin{cases} F_i^{(n)} = \frac{\tilde{F}_i^{(n)}}{\langle \tilde{F}_i^{(n)} \rangle_i} \\ Q_\alpha^{(n)} = \frac{\tilde{Q}_\alpha^{(n)}}{\langle \tilde{Q}_\alpha^{(n)} \rangle_\alpha} \end{cases}. \quad (7)$$

The initial conditions are  $\tilde{Q}_\alpha^{(0)} = 1 \forall \alpha$  and  $\tilde{F}_i^{(0)} = 1 \forall i$ . The vector of country and product scores is the stationary point of these iterative equations. Noteworthy, this algorithm produces highly-nested biadjacency matrices [31].

<sup>7</sup> [https://github.com/COSIN3-UOC/nestedness\\_modularity\\_in-block\\_nestedness\\_analysis](https://github.com/COSIN3-UOC/nestedness_modularity_in-block_nestedness_analysis).

<sup>8</sup> <https://github.com/mat701/BiCM>.

<sup>9</sup> <https://github.com/ganileni/ectools>.

### 3.3.2. PRODY

The PRODY index [32] is the weighted average of per capita GDPs  $Y$ , where the weights represent the RCA<sup>10</sup>  $R_{i\alpha}$  in product  $\alpha$  for country  $i$ :

$$PRODY_{\alpha} = \sum_i \frac{R_{i\alpha} Y_i}{\sum_i R_{i\alpha}}. \quad (8)$$

A slight modification, known as logPRODY and introduced in [33], consists in replacing GDPpc with its logarithm. The reasoning behind this choice is that, since GDPpc's of countries span about four orders of magnitude, the geometric mean is better suited to represent such a numeric distribution of values.

By construction, sectors with high values of (log)PRODY are those where high-income countries play a major role in world exports. Then, under the reasonable assumption that high-income countries display a strong presence where comparative advantages are determined by factors such as know-how, technological skills and so on, sectors characterised by a high (log)PRODY index are more sophisticated than sectors with a low value of the index.

### Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

### Acknowledgments

M S M acknowledges financial support from the URPP Social Networks at the University of Zurich. All the authors acknowledge Dr. Francis Farrelly for data encryption and pre-processing and ISTAT for providing the firm-level data, and in particular Dr Stefano Menghinello and Dr Cristina Lanzi.

### Conflict of interest

The authors declare no competing interests.

### ORCID iDs

Manuel S Mariani  <https://orcid.org/0000-0003-1032-5821>

Andrea Zaccaria  <https://orcid.org/0000-0002-4478-3292>

### References

- [1] Anderson P W 1972 *Science* **177** 393
- [2] Pietronero L 2008 *Europhys. News* **39** 26
- [3] Lynn C W and Bassett D S 2019 *Nat. Rev. Phys.* **1** 318
- [4] Morone F and Makse H A 2019 *Nat. Commun.* **10** 1
- [5] Rohr R P, Saavedra S and Bascompte J 2014 *Science* **345**
- [6] Domínguez-García V and Muñoz M A 2015 *Sci. Rep.* **5** 1
- [7] Morone F, Del Ferraro G and Makse H A 2019 *Nat. Phys.* **15** 95
- [8] Arese Lucini F, Morone F, Tomassone M S and Makse H A 2020 *PLoS One* **15** e0228692
- [9] Lewinsohn T M, Inácio Prado P, Jordano P, Bascompte J and Olesen J M 2006 *Oikos* **113** 174
- [10] Kojaku S and Masuda N 2017 *Phys. Rev. E* **96** 052313
- [11] Solé-Ribalta A, Tessone C J, Mariani M S and Borge-Holthoefer J 2018 *Phys. Rev. E* **97** 062302
- [12] Mariani M S, Ren Z-M, Bascompte J and Tessone C J 2019 *Phys. Rep.* **813** 1
- [13] Flores C O, Valverde S and Weitz J S 2013 *ISME J.* **7** 520
- [14] Lampo A, Palazzi M J, Borge-Holthoefer J and Solé-Ribalta A 2021 arXiv:2105.09191
- [15] Pinheiro R B, Felix G M and Lewinsohn T M 2022 *J. Anim. Ecol.* **91** 2248
- [16] Felix G M, Pinheiro R B P, Poulin R, Krasnov B R and Mello M A R 2022 *Oikos* **2022**
- [17] Morone F 2022 *J. Phys. Complex.* **3** 035007
- [18] Chalas N, Omigie D, Poeppel D and van Wassenhove V 2023 *Isience* **26** 106257
- [19] Palazzi M J, Cabot J, Izquierdo J L C, Solé-Ribalta A and Borge-Holthoefer J 2019 *Sci. Rep.* **9** 1
- [20] Palazzi M J, Solé-Ribalta A, Calleja-Solanas V, Meloni S, Plata C A, Suweis S and Borge-Holthoefer J 2021 *Nat. Commun.* **12** 1
- [21] Serrano M A and Boguná M 2003 *Phys. Rev. E* **68** 015101

<sup>10</sup> The rationale for using revealed comparative advantages as weights is to control for country size when ranking the products.

- [22] Garlaschelli D and Loffredo M I 2005 *Physica A* **355** 138
- [23] Saracco F, Di Clemente R, Gabrielli A and Squartini T 2015a *Sci. Rep.* **5** 1
- [24] Hidalgo C A and Hausmann R 2009 *Proc. Natl Acad. Sci.* **106** 10570
- [25] Tacchella A, Cristelli M, Caldarelli G, Gabrielli A and Pietronero L 2012 *Sci. Rep.* **2** 1
- [26] Cristelli M, Tacchella A, Cader M, Roster K and Pietronero L 2017 *On the Predictability of Growth* (The World Bank)
- [27] Tacchella A, Mazzilli D and Pietronero L 2018 *Nat. Phys.* **14** 861
- [28] Porter M E 1997 *Meas. Bus. Excell.* (<https://doi.org/10.1108/eb025476>)
- [29] Porter M E et al 1998 *Clusters and the new Economics of Competition* vol 76
- [30] Bruno M, Saracco F, Squartini T and Due nas M 2018 *Entropy* **20** 785
- [31] Pugliese E, Zaccaria A and Pietronero L 2016 *Eur. Phys. J. Spec. Top.* **225** 1893
- [32] Hausmann R, Hwang J and Rodrik D 2007 *J. Econ. Growth* **12** 1
- [33] Angelini O, Cristelli M, Zaccaria A, Pietronero L and Podobnik B 2017 *PLoS One* **12** e0177360
- [34] Hausmann R and Hidalgo C A 2011 *J. Econ. Growth* **16** 309
- [35] Cristelli M, Gabrielli A, Tacchella A, Caldarelli G, Pietronero L and Barrat A 2013 *PLoS One* **8** e70726
- [36] Saracco F, Di Clemente R, Gabrielli A, Pietronero L and Preis T 2015b *PLoS One* **10** e0140420
- [37] Mariani M S, Vidmer A, Medo M and Zhang Y-C 2015 *Eur. Phys. J. B* **88** 1
- [38] Miller D J 2004 *Strateg. Manage. J.* **25** 1097
- [39] Jose M L, Nichols L M and Stevens J L 1986 *Financ. Manage.* **33**
- [40] Michel A and Shaked I 1984 *Financ. Manage.* **18**
- [41] Palepu K 1985 *Strateg. Manage. J.* **6** 239
- [42] Pugliese E, Napolitano L, Zaccaria A and Pietronero L 2019 *PLoS One* **14** e0223403
- [43] Christensen H K and Montgomery C A 1981 *Strateg. Manage. J.* **2** 327
- [44] Valvano S and Vannoni D 2003 *Rev. Ind. Organ.* **23** 25
- [45] Kim S H, Jun B and Lee J-D 2021
- [46] Platig J, Castaldi P J, DeMeo D, Quackenbush J and Markowetz F 2016 *PLoS Comput. Biol.* **12** e1005033
- [47] Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 *J. Stat. Mech.* **2008** 10008
- [48] Vinh N X, Epps J and Bailey J 2010 *J. Mach. Learn. Res.* **11** 2837
- [49] Peel L, Larremore D B and Clauset A 2017 *Sci. Adv.* **3** e1602548
- [50] Guimera R, Sales-Pardo M and Amaral L A N 2004 *Phys. Rev. E* **70** 025101
- [51] Payrató-Borras C, Hernández L and Moreno Y 2019 *Phys. Rev. X* **9** 031024
- [52] Bruno M, Saracco F, Garlaschelli D, Tessone C J and Caldarelli G 2020 *Sci. Rep.* **10** 1
- [53] Balassa B 1965 *Manch. Sch.* **33** 99
- [54] Barber M J 2007 *Phys. Rev. E* **76** 066102
- [55] Fortunato S and Barthelemy M 2007 *Proc. Natl Acad. Sci.* **104** 36
- [56] Almeida-Neto M, Guimaraes P, Guimaraes Jr P R, Loyola R D and Ulrich W 2008 *Oikos* **117** 1227
- [57] Duch J and Arenas A 2005 *Phys. Rev. E* **72** 027104
- [58] Squartini T and Garlaschelli D 2011 *New J. Phys.* **13** 083001
- [59] Lancaster H O 1969 *The chi-squared distribution* (New York: Wiley) p 214