Physica A 460 (2016) 254–269

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The mathematics of non-linear metrics for nested networks

Rui-Jie Wu [a], Gui-Yuan Shi [a], Yi-Cheng Zhang [a,b], Manuel Sebastian Mariani [a,*]

[a] Department of Physics, University of Fribourg, 1700 Fribourg, Switzerland
[b] Institute of Fundamental and Frontier Sciences, UESTC, Chengdu 610054, China

HIGHLIGHTS

- We study analytically and numerically the fitness–complexity metric (FCM) and the minimal extremal metric (MEM) for nested networks.
- For both metrics, we derive exact equations for node scores in perfectly nested matrices.
- Our analytic results explain the convergence properties of the fitness–complexity metric.
- In real data, the MEM can produce improved rankings if the input data are reliable.

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form 21 March 2016
Available online 14 May 2016

Keywords:
Nested networks
Non-linear metrics
fitness–complexity algorithm
Economic complexity

ABSTRACT

Numerical analysis of data from international trade and ecological networks has shown that the non-linear fitness–complexity metric is the best candidate to rank nodes by importance in bipartite networks that exhibit a nested structure. Despite its relevance for real networks, the mathematical properties of the metric and its variants remain largely unexplored. Here, we perform an analytic and numeric study of the fitness–complexity metric and a new variant, called minimal extremal metric. We rigorously derive exact expressions for node scores for perfectly nested networks and show that these expressions explain the non-trivial convergence properties of the metrics. A comparison between the fitness–complexity metric and the minimal extremal metric on real data reveals that the latter can produce improved rankings if the input data are reliable.

1. Introduction

Network-based iterative algorithms are being applied to a broad range of problems, such as ranking search results in the WWW [1], predicting the traffic in urban roads [2], recommending the items that an online user might appreciate [3], measuring the competitiveness of countries in world trade [4,5], ranking species according to their importance in plant–pollinator mutualistic networks [6,7], assessing scientific impact [8,9], identifying influential spreaders [10], and many others. While linear algorithms are applied to a broad range of real systems [11,12], it has been recently shown that the non-linear fitness–complexity metric introduced in Ref. [5] markedly outperforms linear metrics in ranking the nodes by their importance in bipartite networks that exhibit a nested architecture [7,13]. The fitness–complexity metric has been originally introduced to rank countries and products in world trade according to their level of competitiveness and quality, respectively [5]. The basic idea of the metric is that while the competitiveness of a country is mostly determined by the diversification of its exports, the quality of a product is mostly determined by the score of the least competitive exporting

* Corresponding author.
E-mail address: manuel.mariani@unifr.ch (M.S. Mariani).

http://dx.doi.org/10.1016/j.physa.2016.05.023

countries. The metric has been shown to be economically well-grounded [5,14], to be highly effective in ranking countries and products by their importance in the network [13], to be informative about the future economic development [15] and the future exports of a country [16]. The metric has been recently applied beyond its original scope and has been shown to be the most efficient method among several network-based methods in ranking species according to their importance in mutualistic ecological networks [7]. In particular, the metric reveals the nested structure of the system much better than the methods used by standard nestedness calculators. Several real systems exhibit a nested structure [5,17–21], which suggests that the metric has a potentially broad range of application.

Despite the relevance of the fitness–complexity metric for nested networks, its mathematical properties and its variants remain largely unexplored. In contrast with linear algorithms such as Google's PageRank [22,11,12] and the method of reflections [23], the convergence properties of the metric cannot be studied through linear algebra techniques. This article provides new insights into the mathematics behind the metric. We study here both the fitness–complexity metric (FCM) and a novel variant, called minimal extremal metric (MEM), that is simpler to be treated analytically. The only input of the metrics is the binary adjacency matrix $\mathcal{M}$ of the underlying bipartite network; we perform here exact computations for perfectly nested matrices, i.e., binary matrices such that a unique border separates all the elements equal to one from the elements equal to zero. For both the MEM and the FCM, we find the exact analytic formulas that relate the ratios of node scores to the shape of the underlying nested matrix. While real nested matrices are not perfectly nested, the expressions derived here for perfectly nested matrices explain the non-trivial convergence properties the metrics found in real matrices [24]. In particular, we analytically determine the condition such that all node scores converge to a nonzero value, which is crucial for the discriminative power of the metrics. This condition has been also found in Ref. [24] (the only previous work that studied the convergence properties of the FCM); differently from the analytic study of Ref. [24] where exact formulas were derived for matrices with two values of node score, in this work we derive by mathematical induction expressions valid for any perfectly nested matrix.

Finally, we contrast the two metrics in real data and show that the MEM can outperform the FCM in packing the adjacency matrix, i.e., ordering its rows and columns in such a way that a sharp curve separates the occupied and empty regions of the matrix [7]. On the other hand, the MEM is more sensitive to noisy data, and, as a consequence, its rankings may be unreliable in the presence of a significant amount of mistakes in the original data [25].

This article is organized as follows: In Section 2, we define the Fitness–Complexity metric (FCM) and the Minimal Extremal Metric (MEM); In Section 3, we analytically compute the ratios between MEM and FCM node scores for perfectly nested matrices and discuss the dependence of the metrics' convergence properties on the shape of the nested matrix; In Section 4, we compare the rankings by the FCM and the MEM in real data of world trade.

2. Non-linear metrics for bipartite networks

In this section, we define the fitness–complexity metric (FCM) and the minimal extremal metric (MEM) for bipartite networks. While the results obtained in this paper hold for any nested matrix, we use here the terminology of economic complexity: rows and columns of the $N \times M$ adjacency matrix $\mathcal{M}$ are referred to as countries and products, respectively; the matrix $\mathcal{M}$ is consequently referred to as the country–product matrix [4]. We label countries by Latin letters ($i = 1, \ldots, N$), products by Greek letters ($\alpha = 1, \ldots, M$); the number of countries and products are denoted by N and M, respectively. The number of products exported by country i and the number of countries that export product α are referred to as the diversification D_i of country i and the ubiquity U_α of product α, respectively [4].

In the fitness–complexity metric (FCM), the fitness scores $\mathbf{F} = \{F_i\}$ of countries and complexity scores $\mathbf{Q} = \{Q_\alpha\}$ of products are defined as the components of the fixed point of the following non-linear map [5]

$$\begin{aligned} \tilde{F}_i^{(n)} &= \sum_\alpha \mathcal{M}_{i\alpha} Q_\alpha^{(n-1)}, \\ \tilde{Q}_\alpha^{(n)} &= \frac{1}{\sum_i \mathcal{M}_{i\alpha} \frac{1}{F_i^{(n-1)}}}, \end{aligned} \tag{1}$$

where scores are normalized after each step n according to

$$\begin{aligned} F_i^{(n)} &= \tilde{F}_i^{(n)} / \overline{F^{(n)}}, \\ Q_\alpha^{(n)} &= \tilde{Q}_\alpha^{(n)} / \overline{Q^{(n)}}, \end{aligned} \tag{2}$$

with the initial condition $F_i^{(0)} = 1$ and $Q_\alpha^{(0)} = 1$.

Eq. (1) implies that the largest contribution to the complexity Q of a product α is given by the fitness of the least-fit exporter of product α. On the other hand, also the fitness scores of the other exporting countries contribute to Q_α; in this sense, the FCM is a quasi-extremal metric [14]. A natural question arises: how would the rankings change when modifying Eq. (1), without changing the main idea behind the metric? A generalized version of the metric where the harmonic terms $1/F$ are replaced by $1/F^\gamma$, with $\gamma > 0$, has been introduced in Ref. [24] and studied in Refs. [24,13]. Here, we introduce a simpler variant of the algorithm, called minimal extremal metric (MEM), where the complexity of a product is equal to

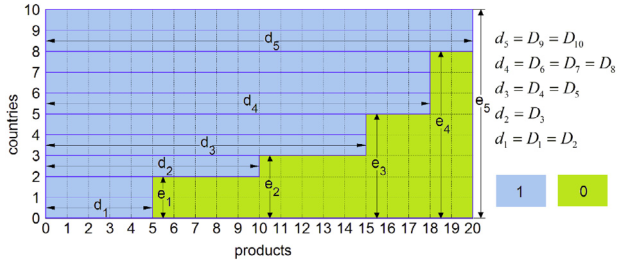

Fig. 1. Illustration of the geometrical meaning of the variables D, d, e in a 10×20 perfectly nested matrix. In this example, there are $m = 5$ groups of countries, which correspond to the diversification values $d_1, \ldots, d_5$, and $m = 5$ groups of products, which correspond to the ubiquity values $N, N - e_1, \ldots, N - e_4$. Due to the perfectly nested structure of the matrix, the groups of countries and products are in one-to-one correspondence: the countries that belong to group i are the least-fit exporters of the products belonging to group i, i.e., of the products whose ubiquity is $N - e_{i-1}$.

the fitness of the least-fit country that exports product α. This metric is extremal, which means that only $\min_{i:\mathcal{M}_{i\alpha}=1}\{F_i^{(n)}\}$ contributes to Q_α. In formulas

$$\begin{aligned} \tilde{F}_i^{(n)} &= \sum_\alpha \mathcal{M}_{i\alpha} Q_\alpha^{(n-1)}, \\ F_i^{(n)} &= \tilde{F}_i^{(n)} / \overline{F^{(n)}}, \\ Q_\alpha^{(n)} &= \min_{i:\mathcal{M}_{i\alpha}=1} \{F_i^{(n)}\}. \end{aligned} \tag{3}$$

Note that the generalized FCM studied in Ref. [13] reduces to the MEM in the limit $\gamma \to \infty$.

3. Analytic results

3.1. Perfectly nested matrix

We focus here on perfectly nested matrices [26], i.e., binary matrices where each country exports all those products that are also exported by the less diversified countries plus a set of additional products. Perfectly nested matrices are also known as stepwise matrices [27], and networks with a perfectly nested adjacency matrix are also referred to as threshold networks [28]. An example of perfectly nested matrix is shown in Fig. 1. In the following, we label countries and products in order of increasing diversification ($D_{i+1} \geq D_i$) and decreasing ubiquity, respectively ($U_{\alpha+1} \leq U_\alpha$). We denote by $\Delta_i := D_i - D_{i-1}$ the number of additional products that are exported by country i but not by country $i-1$, with $\Delta_1 = D_1$.

According to Eqs. (1) and (3), countries with the same level of diversification have the same fitness score, and it is thus convenient to group them together. By doing this, we obtain m groups of countries, with $m \leq N$; we denote by d_i the diversification of countries that belong to group i, where groups are labeled in order of increasing diversification and $i = 1, \ldots, m$. In addition, we denote by e_i the number of countries whose diversification is smaller or equal than d_i. This notation will turn out to be useful for the computations for the FCM. We also define the number $\delta_i := d_i - d_{i-1}$ of additional products that are exported by countries that belong to group i but not by those belonging to group $i-1$, and the number $\epsilon_i := e_i - e_{i-1}$ of countries that belong to group i ($i = 1, \ldots, m$, and $e_0 = d_0 = 0$). Also products are divided into m groups according to their level of ubiquity. Since the number of country and product groups are the same and are equal to m, we use Latin letters ($i = 1, \ldots, m$) to label both groups. Product groups are labeled in order of decreasing ubiquity; we denote by $u_i = N - e_{i-1}$ the ubiquity of the products that belong to group i. The geometrical interpretation of the variables d, D, e is shown in Fig. 1. Note that country and product groups are in one-to-one correspondence: countries that belong to group i are the least-fit exporters of the products that belong to group i.

3.2. Results for the MEM

For a perfectly nested matrix, the fitness of a country $i+1$ at iteration n is given by the fitness of country i at iteration n, plus the complexity of the additional products that are exported by country $i+1$ but not by country i; for the MEM, this property reads

$$\tilde{F}_{i+1}^{(n+1)} = F_i^{(n+1)} + F_{i+1}^{(n)} \times \Delta_{i+1}, \tag{4}$$

where $F_{i+1}^{(n)}$ is the complexity of the additional products. Our aim is to compute the ratios between the fitness scores. We start by considering the two least-fit countries and compute the ratio $F_1^{(n+1)}/F_2^{(n+1)}$ between their scores. Since we are only interested in the ratios between the fitness values, we do not normalize the variables F, Q in the computation; we have then $F_1^{(n)} = \Delta_1^n$ and, starting from Eq. (4)

$$\begin{aligned}\tilde{F}_2^{(n+1)} &= F_1^{(n+1)} + \Delta_2 \times F_2^{(n)} \\ &= F_1^{(n+1)} + \Delta_2 \times (F_1^{(n)} + \Delta_2 \times F_2^{(n-1)}) \\ &= F_1^{(n+1)} + \Delta_2 \times F_1^{(n)} + \Delta_2^2 \times (F_1^{(n-1)} + \Delta_2 \times F_2^{(n-2)}) \\ &= \Delta_1^{n+1} + \Delta_2 \times \Delta_1^n + \Delta_2^2 \times \Delta_1^{n-1} + \cdots + \Delta_2^n \times \Delta_1,\end{aligned} \tag{5}$$

which can be rewritten as

$$\frac{F_1^{(n+1)}}{F_2^{(n+1)}} = \frac{1}{1 + \frac{\Delta_2}{\Delta_1} + \left(\frac{\Delta_2}{\Delta_1}\right)^2 + \cdots + \left(\frac{\Delta_2}{\Delta_1}\right)^n}. \tag{6}$$

If $\Delta_1 = \Delta_2$,

$$\frac{F_1^{(n+1)}}{F_2^{(n+1)}} = \frac{1}{n+1} \underset{n\to\infty}{\longrightarrow} 0 : \tag{7}$$

the ratio converges to zero as $1/n$. The ratio converges to zero also if $\Delta_2 > \Delta_1$, but with an exponential rate:

$$\frac{F_1^{(n+1)}}{F_2^{(n+1)}} \simeq \left(\frac{\Delta_1}{\Delta_2}\right)^n \underset{n\to\infty}{\longrightarrow} 0. \tag{8}$$

By contrast, using the geometric series we can show that the ratio is finite if $\Delta_2 < \Delta_1$:

$$\frac{F_1^{(n+1)}}{F_2^{(n+1)}} \underset{n\to\infty}{\longrightarrow} 1 - \frac{\Delta_2}{\Delta_1}. \tag{9}$$

Interestingly, the three different asymptotic behaviors (7)–(9) correspond to the asymptotic behaviors found in Ref. [24] for the FCM fitness scores with a model matrix where there are only two values F_1 and F_2 of fitness score. We will now use this result as the starting point of a rigorous derivation of the analytic expression for the fitness ratios in an arbitrary perfectly nested matrix.

First, we note that Eqs. (9) and (8) can be summarized as

$$\lim_{n\to\infty} \frac{F_1^{(n)}}{F_2^{(n)}} = 1 - \frac{\Delta_2}{\max\{\Delta_1, \Delta_2\}}. \tag{10}$$

Starting from Eq. (4) and using mathematical induction, we can show that (see Appendix A)

$$\lim_{n\to\infty} \frac{F_i^{(n)}}{F_{i+1}^{(n)}} = 1 - \frac{\Delta_{i+1}}{\max_{j\in[1,i+1]}\{\Delta_j\}}. \tag{11}$$

Note that Eq. (11) relates the score ratios $F_i^{(n)}/F_{i+1}^{(n)}$ to the shape of the perfectly nested matrix, which is encoded in the set of the Δ values. Eq. (11) implies that for any perfectly nested matrix $\mathcal{M}$, all MEM fitness scores converge to a nonzero value if and only if $\Delta_i < \Delta_1 \, \forall \, i > 1$. If the gap Δ_{i+1} between the diversifications of countries i and $i+1$ is the largest gap among the gaps Δ_j of the countries $j \leq i+1$, then the ratio between the score of country i and the score of country $i+1$ converges to zero. The derivation of Eq. (11) is a first example of how the behavior of non-linear metrics can be completely understood for perfectly nested matrices; in the next section, we will derive an analogous expression for the FCM.

Eq. (11) suggests that for a matrix that is not too different from a perfectly nested matrix, the score ratios could be used to assess the convergence of the metric. In particular, one can decide to halt iterations when the following criterion is met:

$$\sum_{i=1}^{N-1} \left| \frac{F_i^{(n)}}{F_{i+1}^{(n)}} - \frac{F_i^{(n+1)}}{F_{i+1}^{(n+1)}} \right| < \epsilon, \tag{12}$$

where ϵ is a predefined accuracy threshold. We refer to Appendix E for the results of the application of this criterion to real data, and to Appendix F for a numerical study of the dependence of the convergence iteration on the size of the system. As first suggested by Ref. [24], if some score ratios converge to zero, countries can be naturally separated in different groups for which all fitness ratios converge to a nonzero value within a set. We refer to Appendix G for a real example from world trade of country separation implied by the existence of zero fitness ratios.

3.3. Results for the FCM

The FCM score of a certain product is determined by the scores of all the exporting countries, which makes the analytical computations for the FCM more difficult than those for the MEM. For the computations with the FCM, it is convenient to group together countries with the same level of diversification. In agreement with the definitions provided in paragraph 3.1, we denote by f_i the fitness of countries that belong to group i, i.e., of those countries whose diversification is equal to d_i. Analogously, we denote by q_i the complexity of the products whose least-fit exporting countries belong to group i. We have then m fitness scores $\{f_1^{(n)}, \ldots, f_m^{(n)}\}$ and m complexity scores $\{q_1^{(n)}, \ldots, q_m^{(n)}\}$. With this notation, we rewrite Eq. (1) as

$$\begin{aligned} \tilde{f}_i^{(n)} &= \sum_{j=1}^{i} \delta_j\, q_j^{(n-1)}, \\ \tilde{q}_i^{(n)} &= \frac{1}{\sum_{j=i}^{m} \epsilon_j / f_j^{(n)}}, \end{aligned} \tag{13}$$

where in the r.h.s. of the second line we replaced $1/f_j^{(n-1)}$ with $1/f_j^{(n)}$, which does not affect the results in the limit $n \to \infty$. Note that in the r.h.s. of the second line, the factor ϵ_j of the terms $1/f_j^{(n)}$ represents the number of countries that belong to group j. Now we transform Eq. (13) into a set of equivalent equations for the fitness ratio $x_i^{(n)} := f_i^{(n)}/f_{i+1}^{(n)}$ and the complexity ratio $y_i^{(n)} := q_i^{(n)}/q_{i+1}^{(n)}$. The equation that relates the scores of two consecutive countries i and $i+1$ is

$$f_{i+1}^{(n+1)} = f_i^{(n+1)} + q_{i+1}^{(n)} \times \delta_{i+1}. \tag{14}$$

In terms of the x variables, Eq. (14) is equivalent to

$$\frac{1}{x_i^{(n)}} = 1 + \frac{\delta_{i+1}\, q_{i+1}^{(n-1)}}{\tilde{f}_{i+1}^{(n)}}, \tag{15}$$

which implies

$$\frac{1/x_i^{(n)} - 1}{1/x_{i-1}^{(n)} - 1} = \frac{\delta_{i+1}}{\delta_i}\, \frac{x_{i-1}^{(n)}}{y_i^{(n-1)}}; \tag{16}$$

reshuffling the terms of this equation, we get

$$x_i^{(n)} = \frac{\delta_i\, y_i^{(n-1)}}{\delta_i\, y_i^{(n-1)} + \delta_{i+1}\,(1 - x_{i-1}^{(n)})}. \tag{17}$$

For the least-fit country ($i = 1$), from Eq. (15) we directly obtain

$$x_1^{(n)} = \frac{\delta_1\, y_1^{(n-1)}}{\delta_1\, y_1^{(n-1)} + \delta_2}. \tag{18}$$

Starting from the second line of Eq. (13) and proceeding in a similar way, we obtain the analogous equations for the y variable:

$$y_i^{(n)} = \frac{\epsilon_{i+1} x_i^{(n)}}{\epsilon_{i+1} x_i^{(n)} + \epsilon_i (1 - y_{i+1}^{(n)})} \tag{19}$$

and

$$y_{m-1}^{(n)} = \frac{\epsilon_m x_{m-1}^{(n)}}{\epsilon_m x_{m-1}^{(n)} + \epsilon_{m-1}}. \tag{20}$$

The set formed by Eqs. (17), (18), (19), (20) is exactly equivalent to the original fitness–complexity equations (Eq. (13)). The uniform initial condition $f_i^{(0)} = 1\, \forall i$ implies the initial conditions

$$x_i^{(0)} = 1, \tag{21}$$

$$y_i^{(0)} = \frac{e_m - e_i}{e_m - e_{i-1}} \tag{22}$$

for the variables x and y. Using Eqs. (17)–(20), we prove the following lemma:

Lemma 1 (*Convergence*)**.** *The sequences $\{x_i^{(n)}\}$ and $\{y_i^{(n)}\}$ are convergent in the limit $n \to \infty$.*

Proof. To prove the convergence, we first prove that the sequences $\{x_i^{(n)}\}$ and $\{y_i^{(n)}\}$ are decreasing in n. From Eq. (18), we have

$$x_1^{(1)} = \frac{\delta_1\, y_0^{(1)}}{\delta_1\, y_0^{(1)} + \delta_2} < 1 = x_1^{(0)}; \tag{23}$$

by combining inequality (23) with Eq. (17), we get $x_2^{(1)} < x_2^{(0)}$; we can repeat the same for all i and get

$$x_i^{(1)} < x_i^{(0)} \quad \forall i. \tag{24}$$

Analogously, by combining inequality (24) with Eq. (20) we get $y_{m-1}^{(1)} < y_{m-1}^{(0)}$, from which we can iteratively show that

$$y_i^{(1)} < y_i^{(0)} \quad \forall i. \tag{25}$$

Now, we use mathematical induction on n to prove that $x_i^{(n+1)} < x_i^{(n)}$ and $y_i^{(n+1)} < y_i^{(n)}$, for all $i = 1, \ldots, m-1$. Suppose that $x_i^{(n)} < x_i^{(n-1)}$ and $y_i^{(n)} < y_i^{(n-1)}$. From Eq. (18), the former inequality directly implies

$$x_1^{(n+1)} = \frac{\delta_1\, y_1^{(n)}}{\delta_1\, y_1^{(n)} + \delta_2} < \frac{\delta_1\, y_1^{(n-1)}}{\delta_1\, y_1^{(n-1)} + \delta_2} = x_1^{(n)}. \tag{26}$$

To prove the inequality $x_i^{(n+1)} < x_i^{(n)}$ for all i, we use mathematical induction on i. To do this, we show that $x_{i-1}^{(n+1)} < x_{i-1}^{(n)}$ implies $x_i^{(n+1)} < x_i^{(n)}$. We obtain

$$x_i^{(n+1)} = \frac{\delta_i\, y_i^{(n)}}{\delta_i\, y_i^{(n)} + \delta_{i+1}\,(1 - x_{i-1}^{(n+1)})} < \frac{\delta_i\, y_i^{(n-1)}}{\delta_i\, y_i^{(n-1)} + \delta_{i+1}\,(1 - x_{i-1}^{(n+1)})} < \frac{\delta_i\, y_i^{(n-1)}}{\delta_i\, y_i^{(n-1)} + \delta_{i+1}\,(1 - x_{i-1}^{(n)})} = x_i, \tag{27}$$

where we used the induction hypothesis on n in the first inequality, and the induction hypothesis on i in the last inequality. A similar proof can be carried out to get $y_i^{(n+1)} < y_i^{(n)}$. Since $x_i^{(n)}$ and $y_i^{(n)}$ are decreasing sequences in n and $x_i^{(n)} > 0$, $y_i^{(n)} > 0$ $\forall n$, then $x_i^{(n)}$ and $y_i^{(n)}$ converge when $n \to \infty$. □

3.3.1. Score ratios when the diagonal does not cross the empty region of $\mathcal{M}$

The lemma ensures the convergence of the non-linear map defined by Eq. (13). We use now the lemma to prove the theorem that guarantees the convergence of the score ratios to a unique fixed point. The theorem holds when the diagonal of the matrix does not cross the empty region of the matrix $\mathcal{M}$, i.e., the region whose elements are zero. In formulas, this condition reads

$$d_i > e_i\, \frac{d_m}{e_m} \quad \forall i = 1, \ldots, m-1. \tag{28}$$

We will also discuss then the procedure to compute the fitness ratios when condition (28) is false. We emphasize that Ref. [24] found this property through an analytic computation on theoretical matrices where two values F_1 and F_2 of fitness score are present, conjectured its validity for any nested matrix and verified this hypothesis through numerical simulations on several datasets. Here, we demonstrate its validity for any perfectly nested matrix.

Theorem 2. *If condition (28) holds, then*

$$\lim_{n\to\infty} \frac{f_i^{(n)}}{f_{i+1}^{(n)}} = a_i, \tag{29}$$

$$\lim_{n\to\infty} \frac{q_i^{(n)}}{q_{i+1}^{(n)}} = b_i, \tag{30}$$

where

$$a_i = \frac{d_i - \frac{d_m}{e_m} e_i}{d_{i+1} - \frac{d_m}{e_m} e_i}, \tag{31}$$

$$b_i = \frac{\frac{e_m}{d_m} d_i - e_i}{\frac{e_m}{d_m} d_i - e_{i-1}}. \tag{32}$$

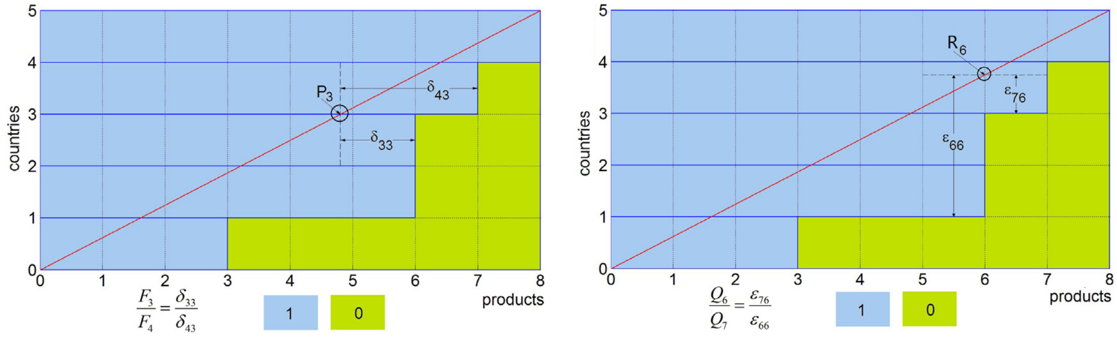

Fig. 2. Illustration of the formulas (36)–(37) for computing the score ratios in a 5 × 8 matrix where the diagonal never crosses the empty region of the matrix. We denote by $\delta_{ij} := D_j - P_{ix}$ the distance between the point P_i where the diagonal of the matrix intersect the line $y = i$ and the line $x = D_j$. We have then $F_3/F_4 = \delta_{33}/\delta_{43}$. Analogously, we denote by $\epsilon_{ij} := R_{iy} - E_j$ the distance between the point R_i where the diagonal of the matrix intersect the line $x = i$ and the line $y = E_j = N - U_j$ We have then $Q_6/Q_7 = \epsilon_{76}/\epsilon_{66}$.

We refer to Appendix B for the details of the proof. The components of the limit vectors (**a**, **b**) have a simple geometrical interpretation. To see this, we rewrite Eqs. (29)–(30) in terms of the original variables F, Q, D, U:

$$\lim_{n\to+\infty} \frac{F_i^{(n)}}{F_{i+1}^{(n)}} = \frac{D_i - \frac{M}{N}i}{D_{i+1} - \frac{M}{N}i}, \tag{33}$$

$$\lim_{n\to+\infty} \frac{Q_i^{(n)}}{Q_{i+1}^{(n)}} = \frac{\frac{N}{M}i - E_{i+1}}{\frac{N}{M}i - E_i}, \tag{34}$$

where we defined $E_i = N - U_i$. In term of the original variables, condition (28) reads

$$D_i > \frac{M}{N}i \tag{35}$$

The solution (33)–(34) has a simple geometric interpretation when considering the representation of the matrix $\mathcal{M}$ in the euclidean plane.

If we denote by P_{ix} the x-coordinate of the point P_i where the diagonal of the matrix — i.e., the diagonal from (0, 0) to (M, N) — intersects the horizontal line $y = i$, we have exactly $P_{ix} = i\,M/N$ (see Fig. 2). As a consequence, assuming $D_i > i\,M/N$ is equivalent to assuming that the diagonal of the matrix never crosses the empty region of the matrix. Eq. (36) can thus be rewritten as

$$\frac{F_i}{F_{i+1}} = \frac{D_i - P_{ix}}{D_{i+1} - P_{ix}}. \tag{36}$$

As shown in Fig. 2, the numerator and the denominator can be interpreted as the distances of the point P_i from the vertical lines $x = D_i$ and $x = D_{i+1}$, respectively. One can also show that condition (35) implies $M\,E_{i+1} < i\,N$ $(i = 1, \ldots, M - 1)$, If we denote by R_{iy} the y-coordinate of the point R_i where the diagonal from (0, 0) to (M, N) intersects the line $x = i$, we have $R_{iy} = i\,N/M$. Eq. (37) can be rewritten as:

$$\frac{Q_i}{Q_{i+1}} = \frac{R_{iy} - E_{i+1}}{R_{iy} - E_i}, \tag{37}$$

which has a simple geometrical interpretation as well (see Fig. 2).

3.3.2. Score ratios when the diagonal does cross the empty region of $\mathcal{M}$

If the diagonal of the matrix crosses the empty region of the matrix — i.e., if there exists some i such that $d_i \leq e_i\,d_m/e_m$ — we cannot directly use Eqs. (29), (30). In this case, the procedure to compute the fitness and complexity ratios is the following:

1. We find the most-fit country j_{max} such that

$$D_i - \frac{i}{j_{max}} D_{j_{max}} > 0 \quad \forall\, i \leq j_{max}. \tag{38}$$

When considering the matrix $\mathcal{M}$ in the euclidean plane, the country j_{max} corresponds to the most-fit country such that the diagonal from (0, 0) to $(d_{j_{max}}, j_{max})$ never crosses the empty region of the reduced matrix that contains only the countries $j < j_{max}$, as shown in Fig. 3.

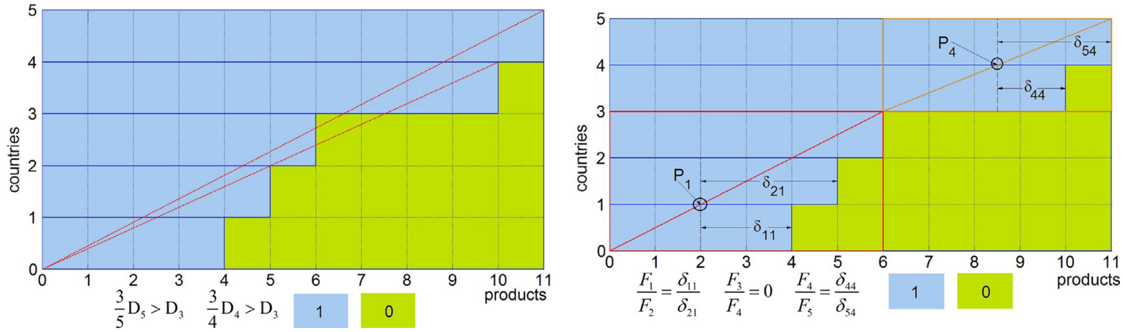

Fig. 3. Illustration of the procedure for computing the score ratios in a 5×8 matrix where the diagonal crosses the empty region of the matrix. In this case, we have to find the most-fit country j_{max} such that condition (38) holds. In this example, $j_{max} \neq 5$ and $j_{max} \neq 4$, because the diagonal from (0, 0) to (d_j, j) crosses the empty region of the matrix for $j = 4, 5$ (left panel). We find $j_{max} = 3$. We can then compute the score ratios for all the countries $i \leq 3$ and all the products $\alpha \leq d_3 = 6$. To do this, we use the same geometrical construction of Fig. 2, but restricted to the submatrix that contains only the three least-fit countries and the d_3 most-ubiquitous products, which corresponds to the block delimited by red border in the right panel. We can then remove from the matrix the countries $i \leq 3$ and the products $\alpha < d_3 = 6$. and compute the score ratios for the countries and products in the residual matrix, which corresponds to the block delimited by orange border in the right panel.

2. Once the value of j_{max} has been determined, we can compute all the fitness ratios for the countries $i < j_{max}$ as

$$\frac{F_i}{F_{i+1}} = \frac{D_i - iD_{j_{max}}/j_{max}}{D_{i+1} - iD_{j_{max}}/j_{max}}, \tag{39}$$

$$\frac{Q_i}{Q_{i+1}} = \frac{ij_{max}/D_{j_{max}} - E_{i+1}}{ij_{max}/D_{j_{max}} - E_i} \tag{40}$$

for all $i < j_{max}$, and $F_{j_{max}}/F_{j_{max}+1} = 0$, $Q_{j_{max}}/Q_{j_{max}+1} = 0$. Note that this formula has the same geometrical interpretation of Eqs. (29)–(30), but the geometrical construction is carried out in a submatrix of the matrix $\mathcal{M}$ (see Fig. 3).

3. We remove from the matrix all the countries $i \leq j_{max}$ and all the products $\alpha \leq d_{j_{max}}$, and restart from point 1, until all the ratios are computed.

The interpretation of this procedure is simple: if the diagonal line crosses the empty region of the matrix, there is at least one pair of countries for which the score ratio converges to zero. In this case, the matrix should be split in blocks such that the score ratios are all nonzero within each block; the score ratios can then be computed inside each block according to Eqs. (39)–(40). A graphical illustration of this procedure is shown in Fig. 3.

4. Results in real networks

4.1. Revealing the nested structure of country–product matrices

The MEM has been introduced in Section 2 as a minimal metric based on the same assumptions of the fitness–complexity metric. In this section, we explore its behavior on real data and compare its rankings with those produced by the FCM. In real data, the fitness–complexity metric has been used to reveal the nested structure of a given network. This is achieved by ordering the rows and the columns of the matrix $\mathcal{M}$ according to their ranking by the metric [5,7]. In particular, the fitness–complexity metric outperforms other existing network centrality metrics and standard nestedness calculators in packing nested matrices [7]. Here, we use the NBER-UN international trade data to compare the matrices produced by the FCM and the MEM; we refer to Appendix C for a detailed description of the dataset. We show here results for the year 1996; results for the different years are in qualitative agreement. We first observe that the rankings of countries by the two metrics are highly correlated ($\rho = 0.994$), and both country scores are highly correlated with country diversification [$\rho = 0.963$ for the FCM, $\rho = 0.955$ for the MEM]. With respect to the matrix produced by the FCM, the matrix produced by the MEM exhibits a sharper border between the empty and the filled parts of the matrix (see Fig. 4). This result suggests that the MEM could be used to produce optimally packed matrices for networks that exhibit a nested structure [7]. In agreement with the convergence criterion introduced in Section 3.2, to obtain the results shown in Fig. 4, we performed 107 and 6700 iterations for the FCM and the MEM, respectively. We refer to Appendices E and F for more details on the convergence properties of the two metrics in real and artificial data.

4.2. Sensitivity with respect to noisy input data

An important issue for any data-driven variable is its stability with respect to perturbations in the system [29,25,30]. Following Refs. [25,13], in order to study the robustness of the rankings with respect to noise, we randomly revert a fraction η of bits in the binary matrix $\mathcal{M}$ and compute the Spearman's correlations of the scores computed before and after the

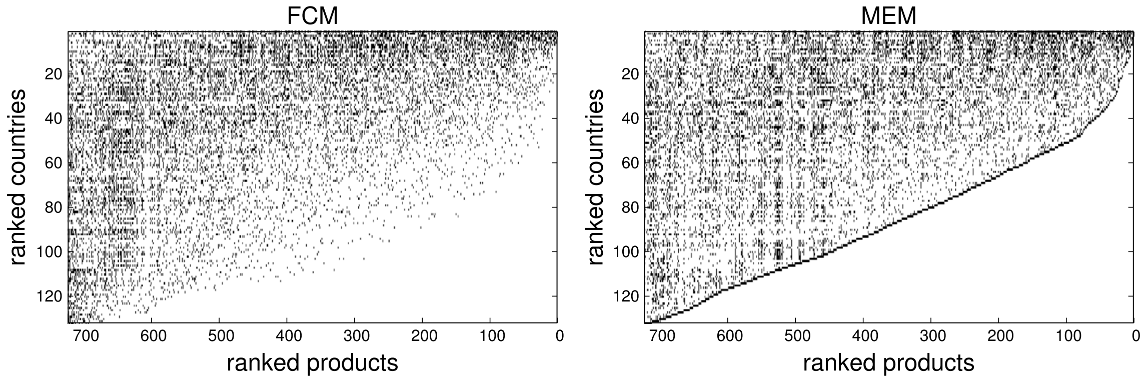

Fig. 4. Country–product matrices resulting from the FCM and the MEM (1996). Both matrices are nested, but the border between the filled and the empty region of the matrix is sharper for the MEM than for the FCM.

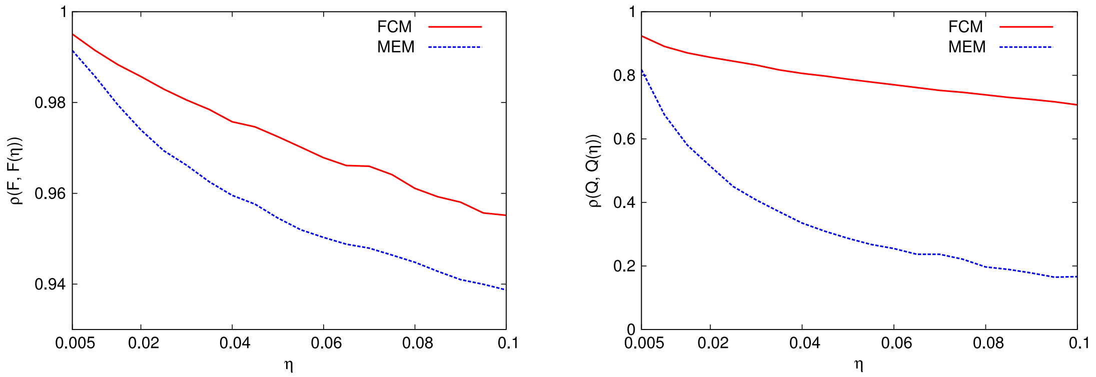

Fig. 5. Robustness against noise of the rankings as a function of the fraction η of reverted bits in the matrix M (year 1996). Robustness is measured by the Spearman's correlation between the rankings before and after the inversion.

reversal. Fig. 5 shows that the rankings by the FCM are more stable than the rankings by the MEM; the gap between the two methods is particularly large for the ranking of products. On the other hand, the different robustness of the two methods is mostly due to the region of the matrix $\mathcal{M}$ whose elements correspond to the most complex products and the least-fit countries. This can be proved by perturbing only the submatrix $\mathcal{M}^{(bottom-right)}$ of $\mathcal{M}$ that contains the $M/2$ most complex products and the $N/2$ least-fit countries according to the ranking by the MEM, and compare the outcome with that obtained when perturbing only the submatrix $\mathcal{M}^{(top-left)}$ that contains the $M/2$ least-complex products and the $N/2$ most fit countries. The difference is striking: for $\mathcal{M}^{(bottom-right)}$, we find $\rho(Q, Q(0.1)) = 0.420$ and $\rho(Q, Q(0.1)) = -0.142$ for the FCM and the MEM, respectively; for $\mathcal{M}^{(top-left)}$, we find $\rho(Q, Q(0.1)) = 0.994$ and $\rho(Q, Q(0.1)) = 0.999$ for the FCM and the MEM, respectively. These findings indicate that the ranking of products by the MEM is not reliable when the data are subject to mistakes and noise, as is the case for world trade data [25], and that the major contribution to the ranking instability comes from the exports of the least-fit countries.

5. Conclusions

Understanding the mathematics behind a network-based ranking algorithm is crucial for its real-world applications. This article moves an important step toward a rigorous understanding of the mathematical properties of the fitness–complexity metric for nested networks. We exactly computed country and product scores for perfectly nested matrices. Our analytic findings are in agreement with the analytic and numeric findings of Ref. [24] on the relation between the convergence properties of the metric and the shape of the underlying nested matrix. We stress again that while we have used the terminology of economic complexity throughout this work, our findings hold for any network that exhibits a nested architecture. For the application of the metric to mutualistic networks, only the meaning of the variables change: F and Q represent active species importance and passive species vulnerability, respectively [7].

In this work, we have also introduced and studied the MEM, a novel variant of the FCM that is simpler to be analytically treated. Our findings on real data indicate that the MEM can order rows and columns of nested matrices even better than the FCM. The high correlation between the country scores obtained with the FCM and the MEM suggests that the MEM and the FCM are similarly informative about the competitiveness of a country in international trade and its future growth potential [15]. On the other hand, the rankings of products by the MEM turn out to be much less stable under a random perturbation in the country–product binary matrix. To conclude, while the MEM can produce more packed nested matrices with respect to those produced by the FCM, its ranking of products is reliable only for high-quality data.

Acknowledgments

This work was supported by the EU FET-Open Grant No. 611272 (project Growthcom), by the Swiss National Science Foundation (Grant No. 200020-143272), by the China Scholarship Council (CSC) scholarship and by the National Natural Science Foundation of China (Grant No. 61503140 and Grant No. 11547040). We wish to thank Alexandre Vidmer for his careful proofreading of the manuscript and his useful suggestions. We acknowledge useful discussions with Matúš Medo, Emanuele Pugliese, Andrea Zaccaria.

Appendix A. Proof of Eq. (11)

We assume that Eq. (11) holds for $i = k$, and show that this assumption implies that it holds also for $i = k+1$. In formulas, we assume that in the limit $n \to \infty$

$$\frac{F_k^{(n)}}{F_{k+1}^{(n)}} = 1 - \frac{\Delta_{k+1}}{H_{k+1}}, \tag{A.1}$$

where $H_{k+1} = \max_{j\in[1,k+1]}\{\Delta_j\}$, and we want to prove that Eq. (A.1) implies

$$\frac{F_{k+1}^{(n)}}{F_{k+2}^{(n)}} = 1 - \frac{\Delta_{k+2}}{H_{k+2}}. \tag{A.2}$$

Using Eq. (4), we obtain

$$\frac{F_{k+2}^{(n+1)}}{F_{k+1}^{(n+1)}} = \frac{F_{k+1}^{(n+1)} + F_{k+2}^{(n)}\,\Delta_{k+2}}{F_{k+1}^{(n+1)}} = 1 + \frac{F_{k+2}^{(n)}\,\Delta_{k+2}}{F_{k+1}^{(n+1)}}. \tag{A.3}$$

We want to express the denominator $F_{k+1}^{(n+1)}$ in terms of $F_{k+1}^{(n)}$ in order to transform this equation into a recurrence relation for $\frac{F_{k+2}^{(n)}}{F_{k+1}^{(n)}}$. To do this, we use Eq. (4) and obtain

$$F_{k+1}^{(n+1)} = F_k^{(n+1)} + F_{k+1}^{(n)}\,\Delta_{k+1} = F_{k+1}^{(n)}\,\Delta_{k+1}\left(1 + \frac{F_k^{(n+1)}}{\Delta_{k+1}\,F_{k+1}^{(n)}}\right) = F_{k+1}^{(n)}\,\Delta_{k+1}\left(1 + \frac{F_k^{(n+1)}}{F_{k+1}^{(n+1)} - F_k^{(n+1)}}\right). \tag{A.4}$$

We now use the hypothesis (A.1):

$$\begin{aligned} F_{k+1}^{(n+1)} &= F_{k+1}^{(n)}\,\Delta_{k+1}\left(1 + \frac{F_k^{(n+1)}/F_{k+1}^{(n+1)}}{1 - F_k^{(n+1)}/F_{k+1}^{(n+1)}}\right) = F_{k+1}^{(n)}\,\Delta_{k+1}\left(1 + \frac{1 - \Delta_{k+1}/H_{k+1}}{\Delta_{k+1}/H_{k+1}}\right) \\ &= F_{k+1}^{(n)}\,\Delta_{k+1}\left(1 + \frac{H_{k+1} - \Delta_{k+1}}{\Delta_{k+1}}\right) = F_{k+1}^{(n)}\,H_{k+1}. \end{aligned} \tag{A.5}$$

Plugging Eq. (A.5) into Eq. (A.3) we get

$$\frac{F_{k+2}^{(n+1)}}{F_{k+1}^{(n+1)}} = 1 + \frac{\Delta_{k+2}}{H_{k+1}}\,\frac{F_{k+2}^{(n)}}{F_{k+1}^{(n)}}, \tag{A.6}$$

Eq. (A.6) is a recurrence equation for $\frac{F_{k+2}^{(n+1)}}{F_{k+1}^{(n+1)}}$. We distinguish two cases:

- If $\frac{\Delta_{k+2}}{H_{k+1}} \geq 1$, then $\lim_{n\to\infty} F_{k+2}^{(n)}/F_{k+1}^{(n)} = \infty$ and $\lim_{n\to\infty} F_{k+1}^{(n)}/F_{k+2}^{(n)} = 0$. In this case $H_{k+2} = \Delta_{k+2}$ by definition and Eq. (A.2) (i.e., the thesis) is satisfied.
- If $\frac{\Delta_{k+2}}{H_{k+1}} < 1$, then we can find the stationary point $\bar{x}$ of Eq. (A.6) by posing $\bar{x} = \frac{F_{k+2}^{(n+1)}}{F_{k+1}^{(n+1)}} = \frac{F_{k+2}^{(n)}}{F_{k+1}^{(n)}}$. We notice that in this case $H_{k+2} = H_{k+1}$ and, as a result, we obtain Eq. (A.2).

This proves the thesis.

Appendix B. Proof of Theorem 2

We denote by $(\mathbf{x}, \mathbf{y})$ the pair of vectors that solve Eqs. (17)–(20) in the limit $n \to \infty$, which read

$$x_i = \frac{\delta_i\, y_i}{\delta_i\, y_i + \delta_{i+1}\,(1 - x_{i-1})}, \quad (i = 2, \dots, m-1) \tag{B.1}$$

$$x_1 = \frac{\delta_1\, y_1}{\delta_1\, y_1 + \delta_2}, \tag{B.2}$$

$$y_i = \frac{\epsilon_{i+1} x_i}{\epsilon_{i+1} x_i + \epsilon_i (1 - y_{i+1})}, \quad (i = 1, \dots, m-2) \tag{B.3}$$

$$y_{m-1} = \frac{\epsilon_m x_{m-1}}{\epsilon_m x_{m-1} + \epsilon_{m-1}}. \tag{B.4}$$

To prove that $(\mathbf{x}, \mathbf{y}) = (\mathbf{a}, \mathbf{b})$ is a solution of Eqs. (B.1)–(B.4), it is sufficient to check that an identity is obtained when replacing x_i and y_i with a_i and b_i in Eqs. (B.1)–(B.4). If $e_m\, d_i > e_i\, d_m$, we can easily use mathematical induction to prove that $x_i^{(n)} > a_i$ and $y_i^{(n)} > b_i$ for all $i = 1, \dots, m-1$ and $n \in \mathbb{N}$. The proof is similar to the proof of Lemma 1. We are then interested only in solutions of Eqs. (B.1)–(B.4) that satisfies

$$a_i \le x_i < 1 \tag{B.5}$$

and

$$b_i \le y_i < \frac{e_m - e_i}{e_m - e_{i-1}}; \tag{B.6}$$

solutions that do not satisfy conditions (B.5) and (B.6) cannot be reached through the iterative process defined by Eqs. (17)–(20), and will not be considered in the following. In the following, always imply conditions (B.5) and (B.6) for the studied solutions. To show that the solution of Eqs. (B.1)–(B.4) is unique, we use a reductio ad absurdum: we assume that a different solution $\mathbf{x} = \tilde{\mathbf{a}}$ and $\mathbf{y} = \tilde{\mathbf{b}}$ exists, and show that this assumption leads to an absurd result. Before doing this, we state two useful properties of the solutions of Eqs. (B.1)–(B.4).

Property 3. *For a solution $(\mathbf{x}, \mathbf{y})$ of Eqs. (B.1)–(B.4), if there exist an integer j such that $x_j > a_j$ or $y_j > b_j$, then $x_i > a_i$ and $y_i > b_i$ for all $i = 1, \dots, m$.*

Proof. Suppose that $x_j > a_j$ for a certain component j. Then

$$y_j = \frac{\epsilon_{j+1}\, x_j}{\epsilon_{j+1}\, x_j + \epsilon_j\,(1 - y_{j+1})} > \frac{\epsilon_{j+1}\, a_j}{\epsilon_{j+1}\, a_j + \epsilon_j\,(1 - y_{j+1})} \ge \frac{\epsilon_{j+1}\, a_j}{\epsilon_{j+1}\, a_j + \epsilon_j\,(1 - b_{j+1})} = b_j. \tag{B.7}$$

In a similar way, one can use Eq. (B.1) to prove the thesis for all $i > j$, and Eq. (B.3) to prove the thesis for all $i < j$. □

In order to have a solution $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ such that $\tilde{a} \neq a$ and $\tilde{b} \neq b$, there must exist at least one component j such that $\tilde{a}_j \neq a_j$ or $\tilde{b}_j \neq b_j$; from the inequalities (B.5)–(B.6) and Property 3, we also have $a_i < \tilde{a}_i < 1$ or $b_i < \tilde{b}_i < (e_m - e_i)/(e_m - e_{i-1})$ for $i = 1, \dots, m$.

Property 4. *If $y_i > 0\, \forall i = 2, 3, \dots, m-1$, for each solution $(\mathbf{x}, \mathbf{y})$ of Eqs. (B.1)–(B.4), the value of y_{m-1} uniquely determines the values of all the other components $\{x_i\}_{i=1}^{m-1}$ and $\{y_i\}_{i=1}^{m-2}$ of the solution. On the other hand, y_{m-1} is uniquely determined by the other components $\{x_i\}_{i=1}^{m-1}$ and $\{y_i\}_{i=1}^{m-2}$ of the solution.*

Proof. The former statement of the Property follows from the fact that if $y_i \neq 0\, \forall i = 2, \dots, m-1$ and we know the last component y_{m-1} of the solution, we can then compute all the other components of the solution and they uniquely depend on y_{m-1}. Suppose indeed that we know the value of y_{m-1}. We can then invert Eq. (B.4) and compute $x_{m-1} = \epsilon_{m-1} y_{m-1} / \epsilon_m (1 - y_{m-1})$, and then plug the obtained x_{m-1} value into Eq. (B.1) to compute x_{m-2}, and then plug the obtained x_{m-2} into Eq. (B.3) to compute y_{m-2}, and so on. The latter statement follows from Property 3 (or equivalently, from the invertibility of all the relations involved in Eqs. (B.1)–(B.4)). □

As a consequence of this property, proving that the solution $(\mathbf{x}, \mathbf{y}) = (\mathbf{a}, \mathbf{b})$ is unique is equivalent to proving that for a solution, the only acceptable value of y_{m-1} is $y_{m-1} = b_{m-1}$.

We will now prove the theorem in two steps:

1. We transform Eqs. (B.1)–(B.4) into a set of equations, hereafter referred to as the transformed equations.
2. We use a reductio ad absurdum and assume that there exists a solution $y = \tilde{b}$ of the original equations such that $\tilde{b}_{N-1} > b_{N-1}$. We use then the transformed equations to show that the solution $y = \tilde{b}$ cannot be a solution of the original equations, which proves the thesis.

B.1. Step 1: Deriving a set of transformed equations

First, we merge the equations (B.1)–(B.4) into two equations

$$x_i = \frac{y_i}{y_i + \frac{\delta_{i+1}}{\delta_i}(1 - x_{i-1})}, \quad (i = 1, \ldots, m-1) \tag{B.8}$$

$$y_i = \frac{x_i}{x_i + \frac{\epsilon_i}{\epsilon_{i+1}}(1 - y_{i+1})}, \quad (i = 1, \ldots, m-1) \tag{B.9}$$

with $x_0 = 0$ and $y_m = 0$. Consider a generic solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ of Eqs. (B.8)–(B.9). Instead of the variables $\{x_1, \ldots, x_{m-1}\}$ and $\{y_1, \ldots, y_{m-1}\}$, we consider the transformed variables $\{x'_1, \ldots, x'_m\}$ and $\{y'_1, \ldots, y'_m\}$ defined by the transformation

$$\begin{aligned} x'_i &= x_{i-1} \quad \text{for } i = 2, \ldots, m; \\ y'_i &= y_{i-1} \quad \text{for } i = 2, \ldots, m. \end{aligned} \tag{B.10}$$

We consider the transformed equations

$$x'_i = \frac{y'_i}{y'_i + \frac{\delta'_{i+1}}{\delta'_i}(1 - x'_{i-1})}, \quad (i = 1, \ldots, m), \tag{B.11}$$

$$y'_i = \frac{x'_i}{x'_i + \frac{\epsilon'_i}{\epsilon'_{i+1}}(1 - y'_{i+1})}, \quad (i = 1, \ldots, m), \tag{B.12}$$

with $x'_0 = 0 = y'_{m+1} = 0$, $\delta'_i = \delta_{i-1}$ and $\epsilon'_i = \epsilon_{i-1}$ for $i = 2, \ldots, m+1$. In the transformed equations, x'_1 and y'_1 are new variables; for a solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ of Eqs, (B.8)–(B.9), the pair of transformed vectors $(\bar{\mathbf{x}}', \bar{\mathbf{y}}')$ satisfies the following set of transformed equations only if $\bar{x}'_1 = \bar{y}'_1 = 0$ for a solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. The values of δ'_1 and ϵ'_1 only affect the values of x'_1 and y'_1, which must be equal to zero for the transformed $(\bar{\mathbf{x}}', \bar{\mathbf{y}}')$ of a solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ of the original equations. This allows us to let δ'_1 and ϵ'_1 be arbitrary parameters in the Eqs. (B.11)–(B.12). Eqs. (B.11)–(B.12) have the same form of Eqs. (B.8)–(B.9). It is possible to show by substitution that a possible solution of Eqs. (B.11)–(B.12) is

$$\bar{x}'_i = \frac{e'_{m+1}\, d'_i - e'_i\, d'_{m+1}}{e'_{m+1}\, d'_{i+1} - e'_i\, d'_{m+1}}, \tag{B.13}$$

$$\bar{y}'_i = \frac{e'_{m+1}\, d'_i - e'_i\, d'_{m+1}}{e'_{m+1}\, d'_i - e'_{i-1}\, d'_{m+1}}, \tag{B.14}$$

where $d'_i = \delta'_1 + \sum_{j=2}^{i} \delta'_j$ and $e'_i = \epsilon'_1 + \sum_{j=2}^{i} \epsilon'_j$. The mth component of this solution is

$$\bar{y}'_m = \frac{e'_{m+1}d'_m - e'_m d'_{m+1}}{e'_{m+1}d'_m - e'_{m-1}d'_{m+1}} = \frac{(\epsilon'_1 + e_m)(\delta'_1 + d_{m-1}) - (\epsilon'_1 + e_{m-1})(\delta'_1 + d_m)}{(\epsilon'_1 + e_m)(\delta'_1 + d_{m-1}) - (\epsilon'_1 + e_{m-2})(\delta'_1 + d_m)}. \tag{B.15}$$

We are interested in the solutions $(\bar{x}', \bar{y}')$ of Eqs. (B.11)–(B.12) such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is solution of Eqs. (B.8)–(B.9), where the transformation $(\bar{\mathbf{x}}', \bar{\mathbf{y}}') \to (\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is given by Eq. (B.10). We can then pose $\bar{y}'_m = \bar{y}_{m-1}$ and $\epsilon'_1 = 1$, and obtain

$$\delta'_1 = \frac{(1 + e_m)\, \delta_m\, (1 - \bar{y}_{m-1})}{\epsilon_m - (\epsilon_m + \epsilon_{m-1})\, \bar{y}_{m-1}} - d_m. \tag{B.16}$$

B.2. Step 2: Reductio ad absurdum

Up to now, we have proven for a solution $(\bar{\mathbf{x}}', \bar{\mathbf{y}}')$ of the Eqs. (B.11)–(B.12) in the form given by Eqs. (B.13)–(B.14), the pair of vectors $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ obtained by the transformation (B.10) is a solution of Eqs. (B.8)–(B.9) if and only if $\bar{x}'_1 = \bar{y}'_1 = 0$, δ'_1 satisfies Eq. (B.16) and $\epsilon'_1 = 1$. We will now show that if we consider a solution of the transformed equations such that $\bar{y}'_m = \tilde{b}'_m = \tilde{b}_{m-1} > b_{m-1}$ and assume that $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ is a solution of the original equations, then the first component $\tilde{a}'_1$ of the solution is different from zero, which is absurd. As a consequence, $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = (\mathbf{a}, \mathbf{b})$ is the only solution of Eqs. (B.8)–(B.9).

Proof. We assume that there exists a solution $\bar{y}_{m-1} = \tilde{b}_{m-1} > b_{m-1}$; from Property 4, all its components $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ are uniquely determined by $\tilde{b}_{m-1}$. Using the solution (B.13)–(B.14) of the transformed equations, from $\tilde{b}_{m-1} > b_{m-1}$ and Eq. (B.16) it follows that $\delta_1' > d_m/e_m$. To prove the thesis, we start by showing that $e_{m+1}' d_i' > e_i' d_{m+1}'$. For $i = 2, 3, \ldots, m$, we have

$$\begin{aligned} e_{m+1}'d_i' - e_i'd_{m+1}' &= (1+e_m)(d_{i-1}+\delta_1') - (1+e_{i-1})(\delta_1'+d_m) \\ &= (1+e_m)d_{i-1} + (e_m - e_{i-1})\delta_1' - (1+e_{i-1})d_m \\ &> (1+e_m)\frac{e_{i-1}}{e_m}d_m + (e_m - e_{i-1})\frac{d_m}{e_m} - (1+e_{i-1})d_m = 0. \end{aligned}$$

For $i = 1$, $e_{m+1}'d_1' = \delta_1' + e_m\delta_1' > \delta_1' + d_m = d_{m+1}' = e_1'd_{m+1}'$, which implies $e_{m+1}'d_i' > e_i'd_{m+1}'$ for all $i = 1, 2, \ldots, m$; as a consequence, $a_1' > 0$, which is absurd. □

Appendix C. The dataset

We use the NBER-UN dataset which has been cleaned and further described in Ref. [31]. We take into account the same list of $N = 132$ countries described in Ref. [32]. For products, we used the same cleaning procedure of Ref. [16]: we removed aggregate product categories and products with zero total export volume for a given year and nonzero total export volume for the previous and the following years. Products and countries with no entries after year 1993 have been removed as well. After the cleaning procedure, the dataset consists of $M = 723$ products. To decide if we consider country i to be an exporter of product α or not, we use the Revealed Comparative Advantage (RCA) [33] which is defined as

$$RCA_{i\alpha} = \frac{e_{i\alpha}}{\sum_{\beta} e_{j\beta}} \Bigg/ \frac{\sum_{j} e_{j\alpha}}{\sum_{j\beta} e_{j\beta}}, \tag{C.1}$$

where $e_{i\alpha}$ is the volume of product α that country i exports measured in thousands of US dollars. RCA characterizes the relative importance of a given export volume of a product by a country in comparison with this product's exports by all other countries. We use the bipartite network representation introduced in Ref. [4], where two kinds of nodes represent countries and products, respectively. All country–product pairs with RCA values above a threshold value – set to 1 here – are consequently joined by links between the corresponding nodes in the bipartite network.

Appendix D. Spearman's correlation coefficient ρ

Given two variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ and $Y = \{Y_1, \ldots, Y_n\}$, we rank them in decreasing order and denote by $\mathbf{x} = \{x_1, \ldots, x_n\}$ and $\mathbf{y} = \{y_1, \ldots, y_n\}$ their corresponding ranking scores. Equal scores are assigned equal ranking positions given by their average ranking position: for instance, if the fourth and the fifth scores in the ranking are equal to each other, then they are both assigned ranking score equal to $(4+5)/2 = 4.5$. The Spearman's correlation coefficient $\rho(\mathbf{X}, \mathbf{Y})$ is then defined as the linear correlation coefficient between the ranking scores, which reads [34]

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{D.1}$$

where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ denotes the mean of $\mathbf{x}$.

Appendix E. Convergence of the metrics in real data

In a perfectly nested matrix, the score ratios converge to a finite value both for the MEM (Eq. (11)) and for the FCM (Eq. (39)). While real matrices are not perfectly nested, one can conjecture that if the matrix is not too sparse, the convergence behavior of a real matrix will be similar. Motivated by this assumption, we define a convergence criterion based on the score ratio, and decide to halt iterations when the following criterion is satisfied

$$d^{(n)} = \sum_{i=1}^{N-1} \left| \frac{F_i^{(n)}}{F_{i+1}^{(n)}} - \frac{F_i^{(n+1)}}{F_{i+1}^{(n+1)}} \right| < \epsilon = 10^{-5}. \tag{E.1}$$

For the country–product matrix shown in Fig. 4, condition (E.1) is satisfied after 107 and 6700 iterations for the FCM and the MEM, respectively (see Fig. E.6). For the FCM, we find that no fitness ratios converge to zero. This is in agreement with our analytic results (see condition (28)) and with the results of Ref. [24], since the diagonal of the matrix never crosses the empty region of the matrix, as shown in the left panel of Fig. 4. For the MEM, we find that three fitness ratios converge to zero,[1] which slows down the convergence of the metric.

[1] We checked that the MEM fitness scores that were converging to zero were still larger than zero after 6700 iterations.

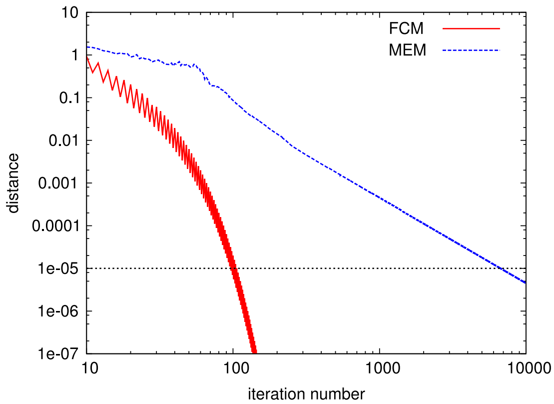

Fig. E.6. $d^{(n)}$ as a function of the iteration number for the country–product matrix represented in Fig. 4.

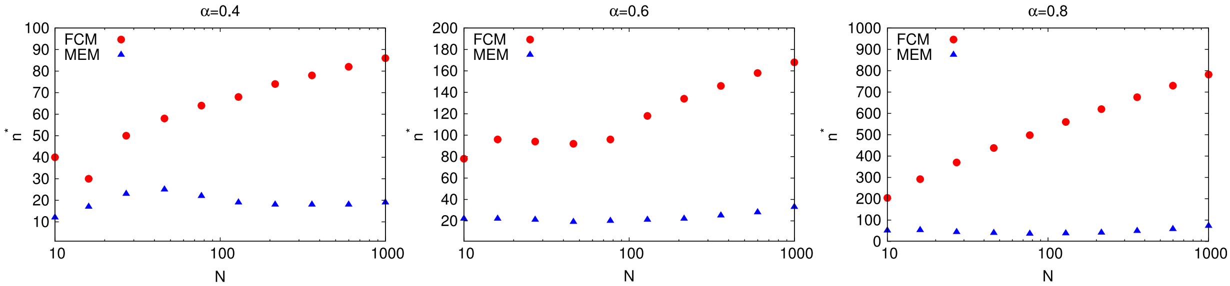

Fig. F.7. Convergence iteration n^* as a function of size N for artificial matrices generated according to Model A described in Appendix F.

Appendix F. Dependence of the convergence iteration of the metrics on network size

In this section, we build artificial nested matrices to study the dependence of the convergence iteration of the metrics on network size. The convergence iteration n^* is defined as the smallest iteration such that condition (E.1) holds. We focus on perfectly nested matrices where the diagonal never crosses the empty region of the matrix, as is the case for the real matrix showed in Fig. 4. We label the countries in order of increasing diversification and the products in order of decreasing ubiquity. To generate the matrices, we use two models:

- *Model A:* This model has a single parameter α which determines the shape of the matrix. In order to have the same ratio M/N as in the real data from world trade analyzed in the main text, we set $M = 5.48\,N$. For country i, we fill the elements corresponding to products $\alpha \in [1, 1 + \lfloor M\, i^{\alpha}/N^{\alpha} \rfloor]$, where α is a parameter of the matrix that determines the shape of the border which separates the empty and the filled regions of the matrix, and $\lfloor x \rfloor$ denotes the largest integer smaller or equal than x. We restrict our analysis to $\alpha < 1$ which corresponds to matrices for which the diagonal does not cross the empty region.
- *Model B:* This model has four parameters x, α, k_1, k_2 which determine the shape of the matrix. Fig. F.8 shows an illustration of a matrix produced with model B. Country 1 has diversification equal to $x + k_1$. For the countries $i \in [2, \lfloor \alpha\,(N-1) \rfloor]$, $d_{i+1} = d_i + k_1$ holds; for the remaining countries, $d_{i+1} = d_i + k_2$ holds. For each value of N, the number of products M is determined by the parameters of the model.

Within this framework, we can study the dependence of the convergence speed of the metrics on network size for a given shape of the matrix's border. For Model A, we find that the convergence speed of the MEM does not strongly depend on system size, as opposed to the convergence speed of the FCM which grows approximately as $\log(N)$ for sufficiently large N (see Fig. F.7). For Model B, we find again a logarithmic growth of the convergence iteration for the FCM for sufficiently large N, whereas the behavior of the MEM can be very different with respect to that found for Model A (see Fig. F.9). In particular, for some parameter settings the convergence of the MEM is slower than that of the FCM, as found in the real data. Figs. F.7 and F.9 indicate that the convergence behavior of the MEM is strongly dependent on the details of the border of the matrix $\mathcal{M}$, as opposed to the FCM which always exhibit asymptotic logarithmic dependence of n^* on N. We did not attempt to investigate the convergence behavior of the metric on alternative matrix models. We envision that suitable modifications of the equations that define the two metrics would mitigate the dependence of convergence speed on system size; however, designing new metrics with improved convergence properties goes beyond the scope of this article.

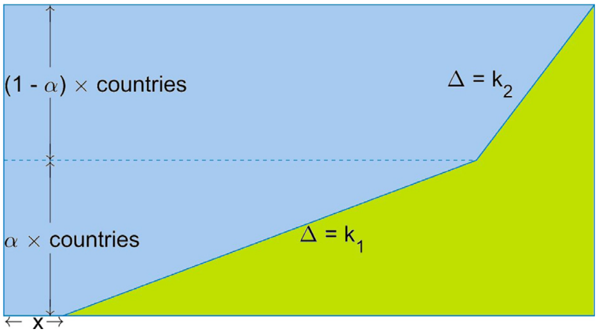

Fig. F.8. An illustration of model B described in Appendix F.

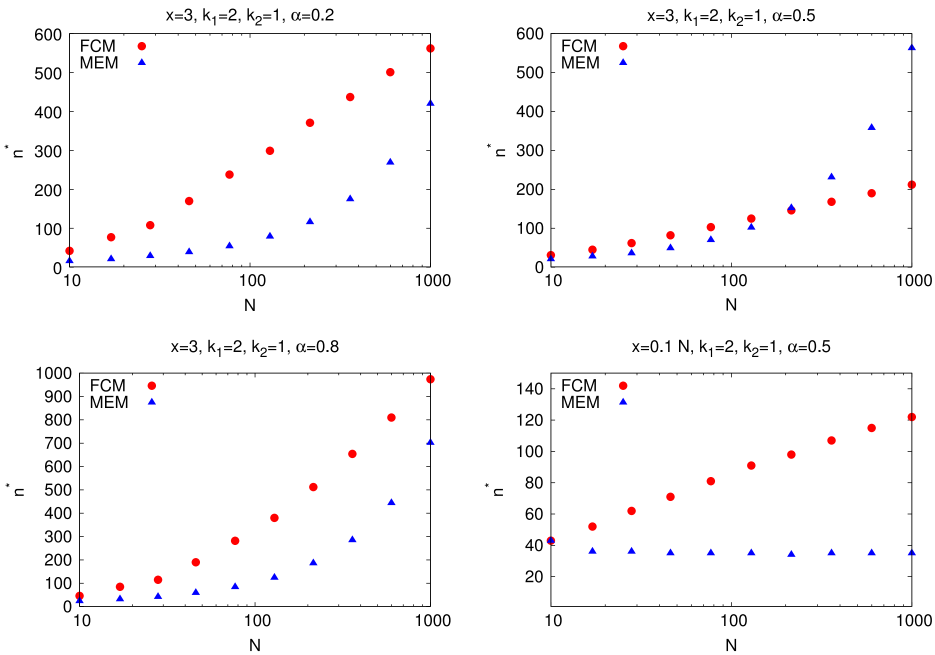

Fig. F.9. Convergence iteration n^* as a function of size N for artificial matrices generated according to Model B described in Appendix F and illustrated in Fig. F.7. The panels show that for the MEM, the dependence of convergence speed on system size strongly depends on the parameters chosen to construct the matrix.

Appendix G. Dividing the matrix $\mathcal{M}$ into submatrices based on fitness ratio

When some fitness score ratios converge to zero, the matrix $\mathcal{M}$ can be separated into different groups of countries such that the score ratios between countries within the same group are always larger than zero. For the FCM, one or more fitness ratios converge to zero when the diagonal of the matrix crosses the empty region of the matrix $\mathcal{M}$ (see Section 3.3.2 and Ref. [24]). For the MEM and for a perfectly nested matrix $\mathcal{M}$, one or more fitness ratios converge to zero when the diversification gap between two countries $i + 1$ and i is equal or larger than the maximum diversification gap of the lower ranked countries, as directly results from Eq. (11). While the criterion for the MEM is not directly applicable to real matrices that are not perfectly nested, we empirically observe in the dataset used for Fig. 4 that the fitness ratios of two pairs of countries converge to zero. As suggested in Ref. [24], we can then separate the countries into three groups such that the fitness ratios are nonzero between any two countries that belong to the same group. The three resulting groups are composed of 103, 2 and 27 countries, respectively (see Fig. G.10, left panel). The right panel of Fig. G.10 shows that the separation of countries into different groups is signaled by discontinuous jumps in the relation between country MEM fitness and country diversification D, which happens for $D < 100$. We emphasize that while the deviation between the trends observed for the FCM and the MEM is relatively small for highly diversified countries, it becomes wide for little diversified countries, which might be relevant for the study of the economic complexity dynamics of developing countries [15].

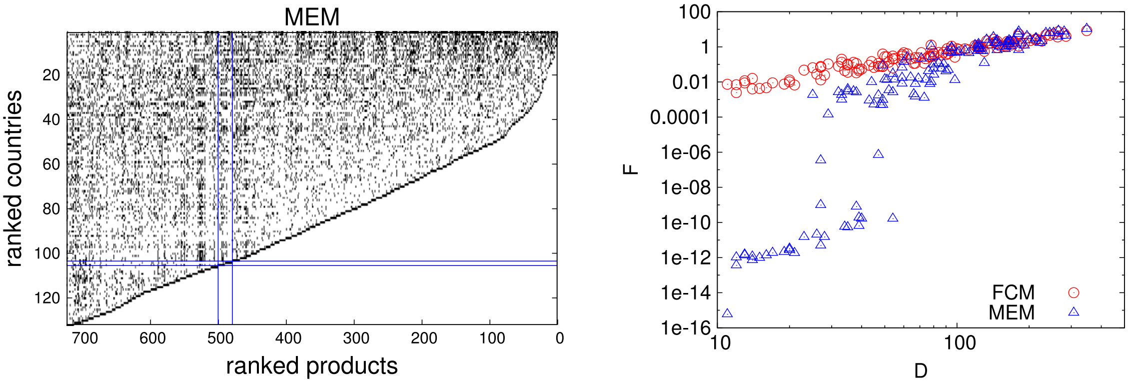

Fig. G.10. *Left:* Country–product matrix resulting from the MEM (1996); horizontal and vertical blue lines separate groups of countries and products, respectively. *Right:* Country fitness score *F* vs. diversification *D* for the FCM and the MEM (1996).

References

[1] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30 (1) (1998) 107–117.
[2] B. Jiang, S. Zhao, J. Yin, Self-organized natural roads for predicting traffic flow: a sensitivity study, J. Stat. Mech. Theory Exp. 2008 (07) (2008) P07008.
[3] L. Lü, M. Medo, C.H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, Recommender systems, Phys. Rep. 519 (1) (2012) 1–49.
[4] C.A. Hidalgo, R. Hausmann, The building blocks of economic complexity, Proc. Natl. Acad. Sci. 106 (26) (2009) 10570–10575.
[5] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, L. Pietronero, A new metrics for countries' fitness and products' complexity, Sci. Rep. 2 (2012).
[6] S. Allesina, M. Pascual, Googling food webs: can an eigenvector measure species' importance for coextinctions? PLoS Comput. Biol. 5 (9) (2009) e1000494.
[7] V. Domínguez-García, M.A. Muñoz, Ranking species in mutualistic networks, Sci. Rep. 5 (2015).
[8] D. Walker, H. Xie, K.-K. Yan, S. Maslov, Ranking scientific publications using a model of network traffic, J. Stat. Mech. Theory Exp. 2007 (06) (2007) P06010.
[9] F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, Diffusion of scientific credits and the ranking of scientists, Phys. Rev. E 80 (5) (2009) 056103.
[10] Z.-M. Ren, A. Zeng, D.-B. Chen, H. Liao, J.-G. Liu, Iterative resource allocation for ranking spreaders in complex networks, Europhys. Lett. 106 (4) (2014) 48005.
[11] D.F. Gleich, Pagerank beyond the web, SIAM Rev. 57 (3) (2015) 321–363.
[12] L. Ermann, K.M. Frahm, D.L. Shepelyansky, Google matrix analysis of directed networks, Rev. Modern Phys. 87 (4) (2015) 1261.
[13] M.S. Mariani, A. Vidmer, M. Medo, Y.-C. Zhang, Measuring economic complexity of countries and products: which metric to use? Eur. Phys. J. B 88 (11) (2015) 1–9.
[14] M. Cristelli, A. Gabrielli, A. Tacchella, G. Caldarelli, L. Pietronero, Measuring the intangibles: A metrics for the economic complexity of countries and products, PLoS One 8 (8) (2013) e70726.
[15] M. Cristelli, A. Tacchella, L. Pietronero, The heterogeneous dynamics of economic complexity, PLoS One 10 (2) (2015) e0117174.
[16] A. Vidmer, A. Zeng, M. Medo, Y.-C. Zhang, Prediction in complex systems: The case of the international trade network, Physica A 436 (2015) 188–199.
[17] J. Bascompte, P. Jordano, C.J. Melián, J.M. Olesen, The nested assembly of plant–animal mutualistic networks, Proc. Natl. Acad. Sci. 100 (16) (2003) 9383–9387.
[18] S. Jonhson, V. Domínguez-García, M.A. Muñoz, Factors determining nestedness in complex networks, PLoS One 8 (9) (2013) e74025.
[19] G. Cimini, A. Gabrielli, F.S. Labini, The scientific competitiveness of nations, PLoS One 9 (12) (2014) e113470.
[20] J. Borge-Holthoefer, R.A. Baños, C. Gracia-Lázaro, Y. Moreno, The nested assembly of collective attention in online social systems, arXiv preprint arXiv:1501.06809.
[21] A. Garas, C. Rozenblat, F. Schweitzer, The network structure of city-firm relations, arXiv preprint arXiv:1512.02859.
[22] P. Berkhin, A survey on pagerank computing, Internet Math. 2 (1) (2005) 73–120.
[23] G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala, A. Tacchella, A network analysis of countries' export flows: firm grounds for the building blocks of the economy, PLoS One 7 (10) (2012) e47278.
[24] E. Pugliese, A. Zaccaria, L. Pietronero, On the convergence of the fitness-complexity algorithm, arXiv preprint arXiv:1410.0249.
[25] F. Battiston, M. Cristelli, A. Tacchella, L. Pietronero, How metrics for economic complexity are affected by noise, Complexity Econ. 1 (1) (2014) 1–22.
[26] W. Ulrich, N.J. Gotelli, Disentangling community patterns of nestedness and species co-occurrence, Oikos 116 (12) (2007) 2053–2061.
[27] M.D. König, C.J. Tessone, Network evolution based on centrality, Phys. Rev. E 84 (5) (2011) 056108.
[28] A. Hagberg, P.J. Swart, D.A. Schult, Designing threshold networks with given structural and dynamical properties, Phys. Rev. E 74 (5) (2006) 056116.
[29] L. Lü, Y.-C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS One 6 (6) (2011) e21202.
[30] J.-G. Liu, L. Hou, X. Pan, Q. Guo, T. Zhou, Stability of similarity measurements for bipartite networks, Sci. Rep. 6 (2016).
[31] R.C. Feenstra, R.E. Lipsey, H. Deng, A.C. Ma, H. Mo, World trade flows: 1962-2000, NBER Working Paper, 11040.
[32] C.A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann, The product space conditions the development of nations, Science 317 (5837) (2007) 482–487.
[33] B. Balassa, Trade liberalisation and revealed comparative advantage1, Manch. Sch. 33 (2) (1965) 99–123.
[34] C. Spearman, The proof and measurement of association between two things, Am. J. Psychol. 15 (1) (1904) 72–101.