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