From Innovation to Diversification: A Simple Competitive Model

Fabio Saracco[1], Riccardo Di Clemente[1] *, Andrea Gabrielli[1,2], Luciano Pietronero[3,4]

1 Istituto dei Sistemi Complessi - ISC CNR UoS "Sapienza" Physics Department Università di Roma, P.le Aldo Moro 5, 00185, Rome, Italy, **2** IMT Institute for Advanced Studies Lucca, Piazza S. Ponziano 6, 55100, Lucca, Italy, **3** Sapienza, Università di Roma, Piazzale A. Moro 5, 00185, Rome, Italy, **4** London Institute for Mathematical Sciences, 35a South St, Mayfair, London, United Kingdom

* riccardo.diclemente@imtlucca.it

Abstract

Few attempts have been proposed in order to describe the statistical features and historical evolution of the export bipartite matrix countries/products. An important standpoint is the introduction of a products network, namely a hierarchical forest of products that models the formation and the evolution of commodities. In the present article, we propose a simple dynamical model where countries compete with each other to acquire the ability to produce and export new products. Countries will have two possibilities to expand their export: innovating, i.e. introducing new goods, namely new nodes in the product networks, or copying the productive process of others, i.e. occupying a node already present in the same network. In this way, the topology of the products network and the country-product matrix evolve simultaneously, driven by the countries push toward innovation.

Introduction

In the economic growth the different, endogenous and exogenous, functional requirements that let a firm pursue products involve the transformation and combination of tangible and intangible attributes [1], such as bureaucratic environment [2], infrastructures [3], education [4], etc. All these features drive either the technologic improvement in the firm production chain [5], or the firm diversification within a country [6], or the introduction of new products. Current models of economic growth consider the relation between the inputs of country goods production and their effects on the overall productivity [7–9], without taking into consideration the measure of the inputs diversity [10].

Economic Complexity, [11–19], is a new expanding field in the economic analysis, which represents a framework to measure the competitiveness of countries and the complexity of products from the national export baskets. The central object of study of this approach is the binary export matrix $\hat{M}$, obtained by imposing a threshold on the Revealed Comparative Advantage (RCA) [20] on the coutry-product trade volumes matrix. The matrix $\hat{M}$ can be thought as the biadjacency matrix of the bipartite network [21–23] in which one layer is represented by countries and the other by exported products.

In order to quantify the competitiveness of countries from the hidden information in $\hat{M}$, a new metric for countries and products has been proposed in [14, 15], overcoming flaws and problems of the seminal work [13]. The basic idea of [15] is to define a non-linear map through an iterative process which couples the Fitness of countries to the Complexity of products. At every step of the iteration, the Fitness F_c of a given country c is proportional to the sum of the exported products, weighted by their complexity parameter Q_p. On the other hand the complexity Q_p of a product p is non linearly related to the fitness of its exporters so that products exported by low fitness countries have a low level of complexity and high complexity products are exported by high fitness countries only.

The historical evolution of $\hat{M}$ shows the development paths followed by the different countries in terms of their export flow. It is possible to build a taxonomy network for products directly from the time evolution of the export baskets of countries [12, 18]. In this way the development pattern followed by different countries can be predicted as the dynamics on an evolving products network.

In this paper we present a dynamical model that describes the evolution of the export baskets of countries by implementing a minimal network model of products innovation processes, which is able to reproduce with good accuracy the main features of the observed evolution of $\hat{M}$. The keystone of our model is the existence of an evolving hierarchical products network in which each country occupies a subset of nodes; within this framework, the products innovation is represented by the introduction of new nodes in the products network. Borrowing the definition from [24] we distinguish between "novelties" and "innovation": "innovation" is something that is new for the whole community, while "novelty" is something known, which is new just for an individual. In this way, the novelty can be "copied" from the near neighbours, while the process that takes to the innovation depends just on the single individual.

There are three main factors that drive the evolution of the export basket of countries and the innovation dynamics of products in our model:

1. the country ability to diversify its basket;

2. the competition within a similar sector of products;

3. the ability to produce innovation with respect to the simple technological updating by adopting already developed technology by other countries. The update of the export basket of a country can take place in two ways: i) as an imitation process from other countries, introducing a novelty, ii) as the development of a brand new product, introducing an innovation. The technological updating is equivalent to the novelties introduction present in [24].

Our model makes the products network and the $\hat{M}$ matrix evolve simultaneously, mutually conditioning one each other: indeed, the country and product that will evolve are chosen on the matrix $\hat{M}$ at that time, but the kind of evolution is decided on the basis of the products network. When the country develops a new product to export, following its path on the products network, it will alter the original $\hat{M}$ matrix by modifying the products network. In this way, the efforts made by countries to develop new technologies modify in real time the path that other countries can take to diversify their own export basket.

The paper is organized as follows. In the section "Methods" we first introduce the ingredients of our model, such as the data set examined (for the comparison of our model with real data) and the network of country and products; then we illustrate in details our algorithm in the subsection "The Model". In "Results" we analyse our results, which are going to be further commented in the section "Discussion".

Materials and Methods

Dataset

The dataset on which we test our model is UN-NBER Sitc Rev2 [25], edited by [26]. From the import registered by the UN, the exports of the World Trade Web (WTW) is reconstructed for nearly 2577 products categories for the years interval 1963-2000. After a data cleaning procedure in order to fix some incoherences, the number of products in the analysed years interval have been fixed to 538, while the number of countries oscillates between 130 and 151, due to geopolitical changes.

The country-product network

Economic Complexity [11–19] focus on the analysis of the bipartite network of countries and exported products. We start from the export volumes matrix $\hat{q}$: every entry q_{cp} represents the total amount of exports in USD of the product p by the country c. In order to binarize q, the RCA (*Revealed Comparative Advantage*, [20]) is calculated:

$$\text{RCA}_{cp} \overset{\text{def}}{=} \frac{\dfrac{q_{cp}}{\sum_{p'} q_{cp'}}}{\dfrac{\sum_{c'} q_{c'p}}{\sum_{c',\,p'} q_{c'p'}}}. \tag{1}$$

The philosophy at the basis of Eq (1) is to give a non dimensional measure of how the export basket of a specific country is organized respect to the average, comparing the impact of the product p on the the export basket of c respect to the impact of p on the global export basket. In the light of that, we can impose a threshold on the RCA-matrix, obtaining the binary $\hat{M}$-matrix: if m_{cp} is the entry for the $\hat{M}$ relative to the country c and the product p, then

$$m_{cp} = \begin{cases} \text{if } \text{RCA}_{cp} \geq 1 & 1 \\ \text{if } \text{RCA}_{cp} < 1 & 0 \end{cases}, $$

i.e. only exported products exceeding the RCA threshold appears in the basket of a country. An export basket in which just raw materials are over the threshold of the RCA (so that appear in the $\hat{M}$ matrix) denotes limited industrialization, while a diversified one, from highly exclusive products to most simple ones, implies a completed industrialization. The matrix $\hat{M}$ can be thought as the biadjacency matrix of a bipartite network, in which the first layer, corresponding to the row index c, is composed by countries, while the second layer, corresponding to the column index p, is composed by the products. Links are permitted only between nodes of different layers.

Traditionally [27], the degree of the nodes, i.e. the number of links per node, are called *diversification* for countries and *ubiquity* for products; in terms of $\hat{M}$ they can be respectively expressed as

$$k_c \overset{\text{def}}{=} \sum_p m_{cp}; \qquad k_p \overset{\text{def}}{=} \sum_c m_{cp}. \tag{2}$$

The model

Our model focuses on the historical evolution of the $\hat{M}$-matrix, i.e. the biadjacency matrix relative to the bipartite undirected binary network of countries and exported products. The evolution of $\hat{M}$ is driven by the evolution of the products network [12, 18], i.e. a hierarchical

network based on the productivity processes such that two different products are linked if there is the possibility of passing from one to the other by a technological improvement. The product network takes the topology of a forest in which the "roots" represent the ancestors product, like raw materials, while the most outer leaves are the highest technology goods.

The tree-like topology may appear as a great simplification, in the sense that a certain production could be affected even by a "distant" technological improvement. Anyway, the topology proposed has been shown, [12, 18], to be a reliable tool able to capture the main features of countries productivity evolution; it is indeed remarkably that a so simple structure can correctly reproduce the evolution of countries diversification. A pictorial representation of the products network can be found in the left part of the top panel of Fig 1: the network nodes represent different products and the links are the technological relationship, while colored disks occupying a given node stand for the different countries (one color for each country) able to export the given product.

We superimposed the information contained in matrix $\hat{M}$ on the topology of the products network in order to give a more immediate interpretation of the mechanism of the export baskets evolution we are proposing.

Let us recall some definitions from [24]: in this inspiring paper focused on the nature of innovation, the authors distinguish between "novelties" and "innovations". A novelty is a tool, a webpage, a song or any item you can think about that is relatively new, literally already present in the common knowledge, but not already experienced by the single agent (a person or, as in the present article, a country); in contrast, an innovation is something that has never appeared in the set of known items, so it is new for everyone.

In our model, a country develops its export basket progressively occupying a subset of the product network. More specifically, at each time step of the algorithm, a selected country either occupies nodes already occupied by other country or creates a new node by sprouting in the product network *nearby* one of those already present in its export basket. Considering as the set of nodes that a country can occupy at a given time steps just the closest products its own export basket is similar to Kauffman conjecture, [28], about the "adjacent possible" nature of evolution: Kauffman proposed that the innovation process takes place only on the border of one's own set of knowledge as items close to the borders are the most probable to be investigated and introduced.

Our algorithm is implemented as the sequential iteration of three fundamental substeps at each time step: the first one decides the country that will enlarge its basket; the second substep selects the product that will drive the evolution on the product network; finally, the third one will decide the path of the country basket evolution, either creating a brand new product (thus, innovating), or copying a product by adopting already developed technology by other countries (thus, introducing a novelty).

In details, the 3 sequential substeps of our model are the following:

1. **The country selection: Divesification**. At the first substep, a country c has a probability to be selected

$$P_1(c) \sim k_c^\alpha \tag{3}$$

where k_c is its diversification, as defined in Eq (2), and $\alpha > 0$ is a parameter of the model. Normalization of the all countries is imposed to Eq (3) in order to evolve a single country at each time step. This first substep is similar to the generalization of the preferential attachment presented in [29], with the difference that here we select a node on one layer of a bipartite network, while in [29] a node was selected in a monopartite one.

Eq (3) says that countries with a diversified export basket have a higher probability to be
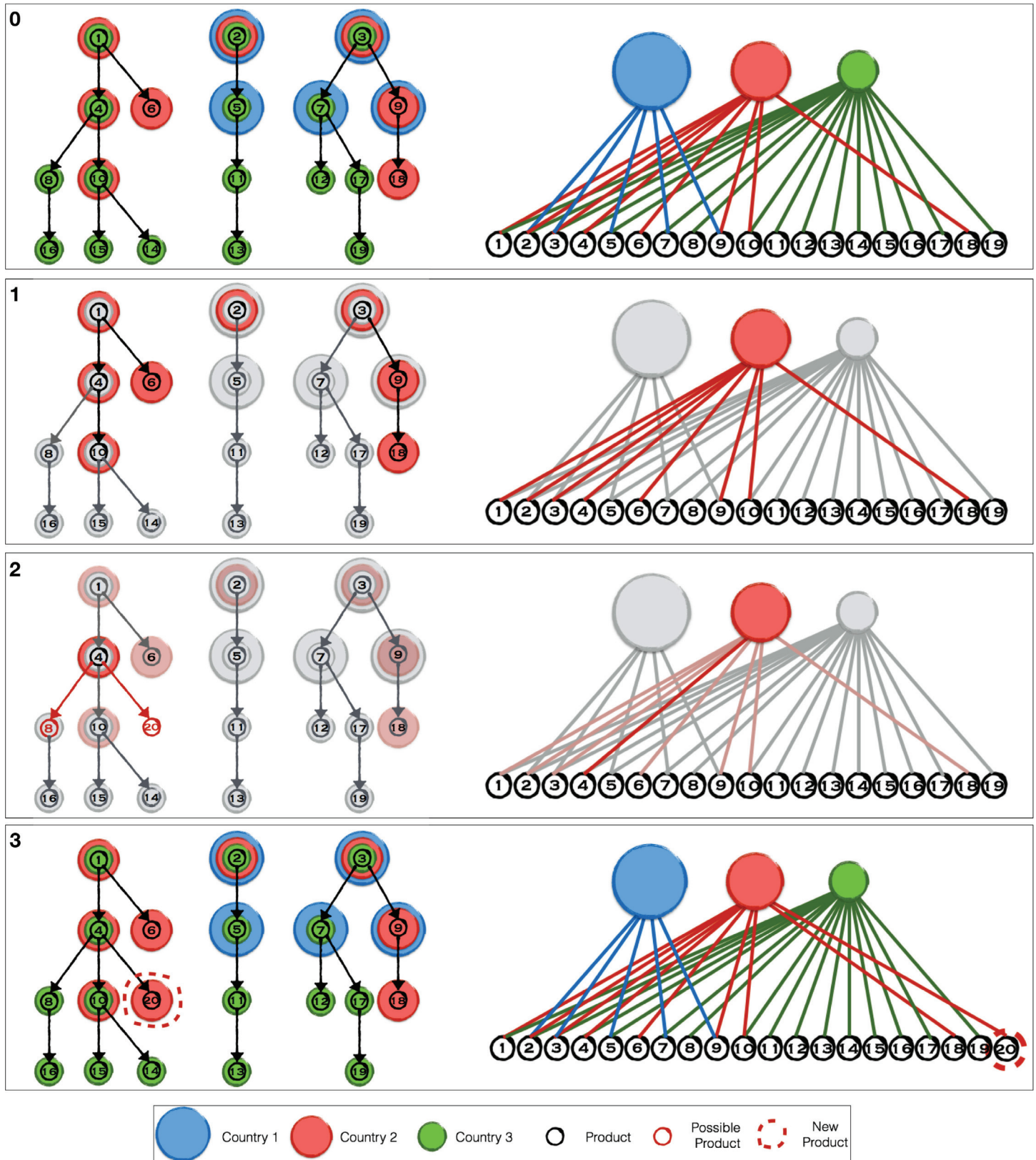
Fig 1. Model Evolution. We propose a pictorial example of one iteration of the evolution of our model: the bipartite network of countries and exported product is on the right, while the the products network is on the left. Coloured (red, blue and green) disk represent countries, while products are circles

numbered from 1 to 20; in the bipartite network the link are coloured as the country they refer to. For completeness, we projected the information from the bipartite network on the products network, superimposing coloured disk on a given node if the country is able to export the single product. In the top panel of Fig 1 we present the initial condition, in which the red, blue and green countries have their export products and occupy different nodes in the products network. In the second panel of Fig 1 the first substep of the algorithm is taken: following the recipe described in the section "The method" the red country is selected. The products in "red" export basket are highlighted both on the bipartite network (on the right) and on the products network (on the left). At the second substep, among the products in the "red" export basket, "4" is selected in the third panel of Fig 1. Together with the previous step, we have selected a link in the bipartite network. On the left, the possible choices on the products network for the third substep: since "10" is already in the "red" export basket, it cannot be selected as the final target for the evolution, so the only possibility left are product "8" (already produced by "green") or "20", a brand new product. In the bottom panel of Fig 1 the thirds choice has been taken and "20" is a new product in the products network.

chosen: it implements the idea that the diversification can be taken as a good proxy for efforts a country makes in order to evolve its export basket (a wider discussion is developed in [30]). The selected country is the one performing the evolution of matrix $\hat{M}$ in the next two substeps.

In the second panel of Fig 1 the first step is pictorial shown: among the three countries (red, blue and green), represented by different disks in the upper layer of the bipartite network, the red one is selected and its export basket in the product network highlighted. Note that the diversification k_c is the number of nodes occupied by each country, so we have $k_{\text{red}} = 8$, $k_{blue} = 5$ and $k_{\text{green}} = 16$.

2. **The evolving product selection: Competition**. Once the country c has been chosen, we have to select a product already in its export basket, from which either moving towards an unoccupied existing neighbouring nodes or sprouting a brand new one. We select such a product p with a probability:

$$P_2(p|c) \sim k_p^\beta \qquad (4)$$

where k_p is the ubiquity, as defined in Eq (2) and $\beta > 0$ is the second parameter of the model. Similarly to Eqs (3) and (4) implements the generalization of the preferential attachment criterion of [29], applied to the product layer of our bipartite network.

The idea is that the more producers, the harder the efforts on renovation, the more possibilities of improving the export basket may come from the most ubiquitous products. In effect, similar behaviours has been shown by experimental evidences, [31].

The third panel of Fig 1 illustrates such second substep: from the commodities present in the "red" country export basket, $p =$ "4" is selected (while on the bipartite network k_p is the number of countries linked to p, in the products network it is represented by the number of differently coloured disk occupying the selected node).

3. **Target product selection: Innovation against Novelty**. In the previous substeps, we have selected the country c and the product p (in the pictorial representation of Fig 1 the "red" country and the product "4") performing the last evolution substep. This choice was based on the properties of the matrix $\hat{M}$ at that time steps; more precisely on diversification and ubiquity defined in Eq (2). So far, no information from the products network topology has been used.

Let us now consider the position the chosen p occupies in the products network. There are two options: either introducing a new node in the products network, i.e. innovating, or evolving along the links already present in the products network by introducing in the export basket a product neighbour of p already exported by other countries, i.e. introducing a novelty.

The probability of copying novelties from others will increase with the number of countries that already export the given good. In this way we implement the idea that it is much easier to acquire close technology. On the other hand, the innovation process of proposing brand new products will need an extra parameter.

As possible novelties we consider all the first neighbours p' of p in the product network which are not already present in the basket of c. At the same time let us call p^* a possible brand new product sprouting out of p. We call $\tilde{p}$ the generic element of the set obtained by the union of p' and p^*. In this third substep we select a single element $\tilde{p}$ of this set with the probability given by:

$$P_3(\tilde{p}|c,p) \sim (k_{\tilde{p}} + k^0)^\gamma, \tag{5}$$

where $\gamma, k^0 > 0$ are the last two parameters of the model and $k_{\tilde{p}}$ is the ubiquity of $\tilde{p}$; clearly $k_{p^*} = 0$. The quantity $k^0 > 0$ is a necessary an offset permitting even the artificial product p^* to be selected. The $k_{\tilde{p}}$ term, makes the probability of "copying" other "accessible" products made already by other country larger than introducing an innovation.

The third panel of Fig 1 illustrates this third substep: in the second panel of the same Fig 1 we selected the product p = "4".

Now the possibilities are between either selecting the product "8", already produced by the "green" country, or introducing a brand new product p^* = "20". In the picture we represent pictorially this last event.

In our algorithm we iterate these three substeps until the number of products in the networks is the same of the observed matrix we want to reproduce, i.e. 538 for the examined data. Summarizing the parameters of the model are $\alpha, \beta, \gamma, k^0$.

The density saturation. For all the simulated values of the parameters once the number of products introduced in the network reaches the same number of the observed bipartite network, the density of links is however smaller than the observed one. Consequently starting from this time-step we will set $k^0 = 0$ in order to prevent the creation of new products and permit only the introduction of novelties increasing in this way the density of bipartite network. We stop the iteration of the algorithm when the density of the country-product bipartite network saturates to the observed one is $\rho_{cr} \simeq 0.13$. Interestingly we observed that a similar behaviour is shown by real data as illustrated by Fig 2. In this figure we can appreciate that the density of the bipartite network increases from year to year up to 1975 and then saturate to an approximate constant values. One possible explanation of this phenomenon is probably the fact that products categories of Sitc Rev.2 have been formalized in 1980 and data for the years before 1980 have been converted to the Sitc Rev. 2 by the Sitc Rev.1 and some data may have been lost in merging the datasets.

Initial Conditions. In order to render the final results of our model less sensible to the initial conditions, we compose the initial exports basket of each country as a Bernoulli trial. We start with a given number N_{roots} of initial products, hereby called *roots*. We assign each product of this set to each country with a constant probability P_0, independently of the other countries. Consequently each country will have a random number of initial products selected from the set of roots, whose probability distribution is binomial with mean value $N_{\text{roots}}P_0$ and standard deviation $\sqrt{N_{\text{roots}}P_0(1 - P_0)}$. At this level the product network is a set of disconnected nodes, represented by the roots; at the first step of the evolution dynamics the product network will evolve from this initial condition as a branching process in which the branching event is represented by innovation.
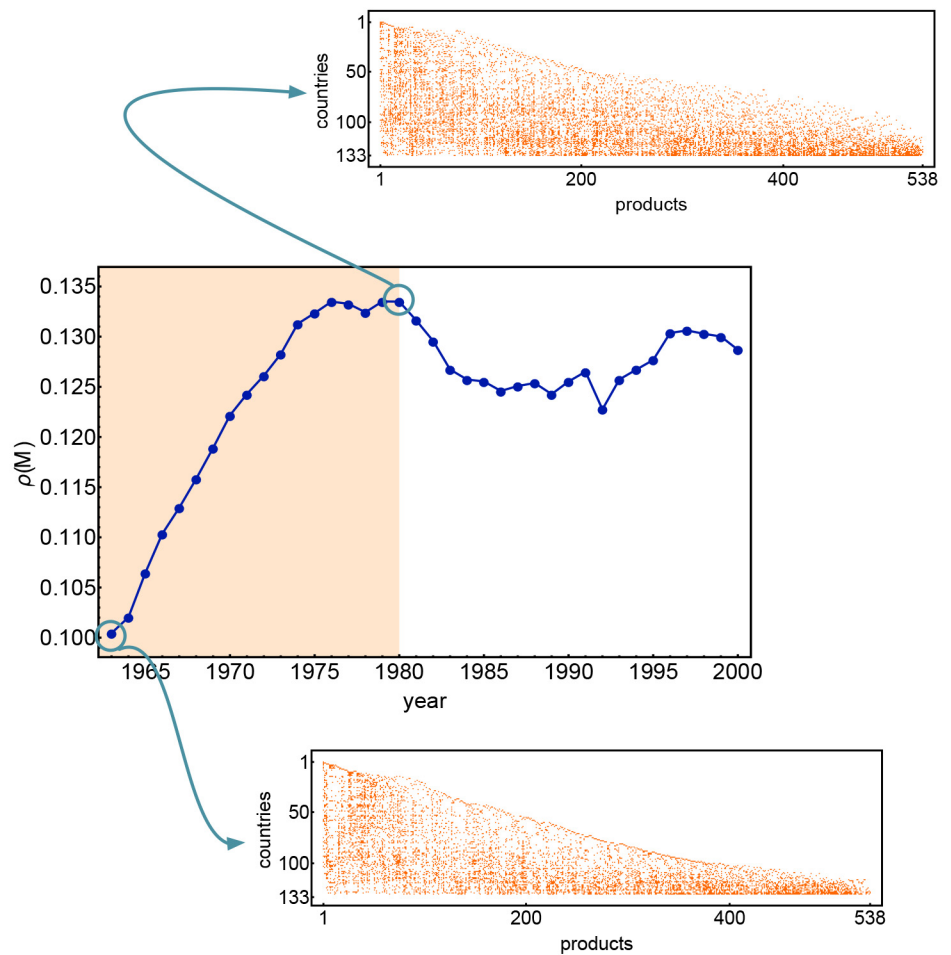
Fig 2. The density evolution for the dataset [25, 26]. It is possible to observe the density increasing until a certain value $\rho_{cr} \sim 0.13$ for the year ~ 1975; our model follows a similar behaviour, limiting the evolution to the export of existing products once the number of nodes in the products network reaches the number of observed products in the real network.

Exploring parameter space. As explained in the Supporting Information in S1 File, in order to find the best range for the values of the parameters characterizing both the initial condition, (N_{roots}, P_0), and the evolutionary model, $(\alpha, \beta, \gamma, k^0)$, we tested for each choice of the first pair, in the range of reasonable values of N_{roots} from 10 to 40 and of P_0 from 0.2 to 0.6, a wide range of the dynamical parameters. We found that the performance of the algorithm weakly depends on the precise choice of $(\alpha, \beta, \gamma, k^0)$ around reasonable values for the exponents α, β, γ around the unitary value, and for k^0 around the minimal observed value $k^0 = 4$ for the product ubiquity k_p in real data. The slightly best performing set of values is however found to be $N_{roots} = 20$, $P_0 = 0.3$, $\alpha = 1.55$, $\beta = 0.8$, $\gamma = 0.3$, $k^0 = 4$.

Results

We compare the matrices $\hat{M}$ generated by our algorithm for different values of the parameters $(\alpha, \beta, \gamma, k^0)$ with the observed one using several non-trivial quantities characterizing the

features of binary bipartite networks. In particular we will use the countries Fitnesses and products Complexities distributions, as defined in [14, 15]; nestedness [32]; assortativity [33]; motifs for bipartite networks [34–36].

Fitness and Complexity In spite of the simplicity of our the model, there is a remarkably good agreement between simulations and real data for Fitness/Complexity (for details about the definition of Fitness and Complexity, see the Supporting Information in S1 File). In particular the shape of the scatter plot for countries fitnesses (products complexities) ranking against countries diversifications (products ubiquities) reproduces well the behavior in the real data; the result is shown in Fig 3(a) (Fig 3(b)). It is possible to see that our algorithm is able to reproduce the "shape" of the original matrix data (blue dots) within the 95%, which is a remarkable result, since these peculiar trends are derived by the highly non-linear algorithm for Fitnesses and Complexities.

Nestedness The concept of nestedness relates to how much a row (or a column) is subset of the others. In this sense, it is a way to evaluate the "triangularity" of the binary matrix $\hat{M}$. Among the different definition of nestedness [32, 37–39], we opted for the NODF (*Nestedness measure based on Overlap and Decreasing Fill*), presented in [32]. This choice is motivated by the fact that is independent form the order of the elements and it is particularly intuitive. The final value of nestedness is the sum of the contributions by the columns (i.e. products) and the rows (i.e. countries), weighted by the possible couple of elements (for details about the definition of the NODF, see the Supporting Information in S1 File); for completeness, we analysed the contributions from countries and products separately, i.e. NODF_c and NODF_p respectively, as well as the total contribution, NODF_t.

As it can be seen in the Fig 4, panels c, the NODF_p of real data is well replicated by our algorithm. This means that our model is able to catch the main features of the hierarchical organization of products; this result is due to the assumption of the presence of a products network, i.e. a hierarchical structure among products. The result for NODF_c, in the panel b of Fig 4, is much more non trivial: in this case the values from the simulations reproduce the same quantity for the real matrix, even though no explicit structure was imposed on the set of countries.

As a matter of fact countries, by following productivity paths on the products network, impose a nested structure to countries as a consequence of the i) hierarchical structures of products, and ii) the mechanism of the products network evolution. Since poorly diversified countries cover a subset of the products network of the most diversified one, the matrix $\hat{M}$ appears nested respect to rows too.

Assortativity As proposed in [33], assortatitvity is a measure of how much nodes link nodes with similar degree (more details can be found in the Supplementary information). Since the poorly diversified countries focus their export toward the most ubiquitous products, the export bipartite network is disassortative, i.e. a network in which low degree nodes links to high degree ones. Our model provides a value inside the 95% of probability, but, respect to other measurements, the assortativity shows a worse agreement between the simulations and real data, as can be seen in Fig 4(d). Actually, this behaviour is probably due to a sort of "second order" effect: the most diversified countries do exports even the most ubiquitous products, but their export basket is nevertheless biased towards the highest quality products. In effect, the distribution in Fig 4(d) shows that the model provides a less disassortative bipartite network respect to the real one (in red in the figure).
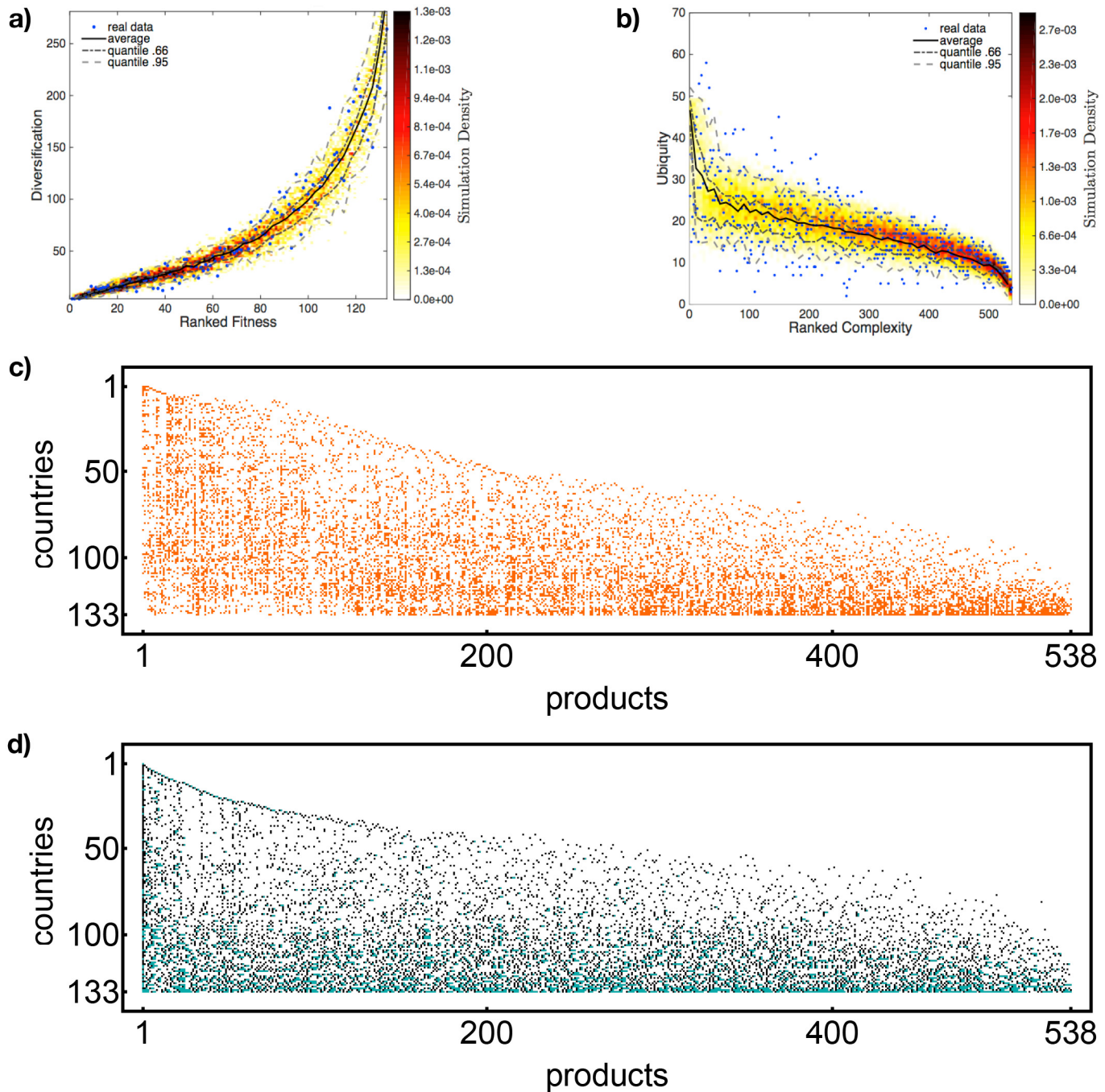
Fig 3. Model Results. In **(a)** the scatter plot of Fitness ranking against countries diversification, while in **(b)** the one for Quality ranking against products ubiquity; the blue points represent the observed values (for the year 1980 from the dataset of [25, 26]). The black line represents the average value on the simulations, while the grey lines bind the area between the second and the first 3-quantiles (dot-dashed) and between the 975th and 25th permilles (dashed). The data obtained are for initial conditions $N_{roots} = 20$ and $P_0 = 0.3$ and parameters $\alpha = 1.55$, $\beta = 0.8$, $\gamma = 0.3$, $k^0 = 4$. In the $\sim 82\%$ the observed data fall into the area between 975th and 25th permilles for the fitness distribution, $\sim 75\%$ for the quality distribution. In **(c)** the original matrix for 1980 from the dataset of [25, 26]; in **(d)** one of the synthetic matrix for initial conditions $N_{roots} = 20$ and $P_0 = 0.3$ and parameters $\alpha = 1.65$, $\beta = 1.1$, $\gamma = 0.6$, $k^0 = 4$.

doi:10.1371/journal.pone.0140420.g003
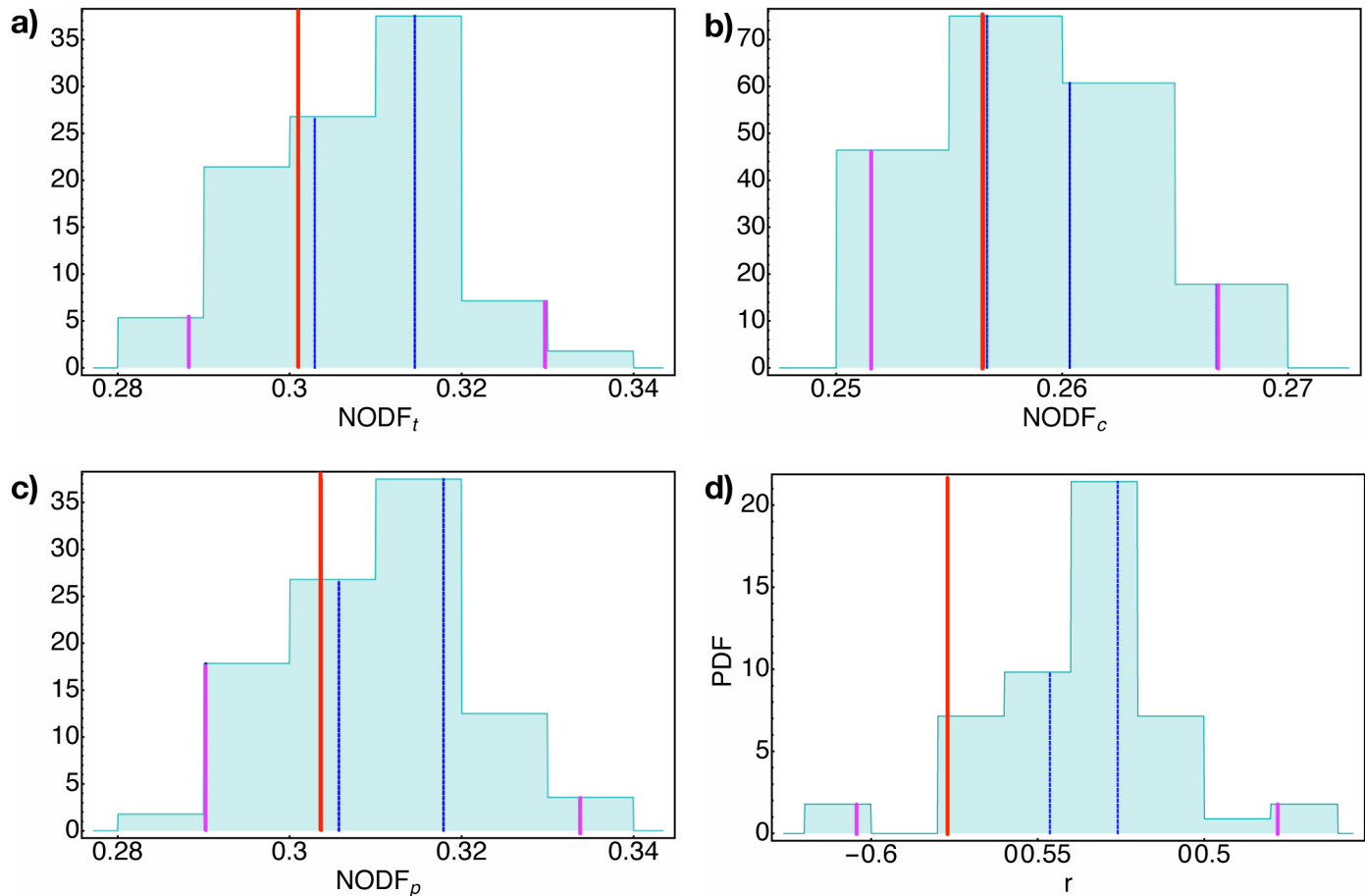
Fig 4. Nestedness & Assortativity. The distributions for the nestedness values (obtained employing NODF, the definition by [32]) and assortativity index r (obtained employing the definition by [33]) for 50 simulations with initial conditions $N_{roots} = 20$ and $P_0 = 0.3$ and parameters $\alpha = 1.55$, $\beta = 0.8$, $\gamma = 0.3$, $k^0 = 4$. In **a)**) the total NODF, in **b)** the NODF for rows and in **c)** the one for columns. The red line is the observed value for the year 1980 from the dataset of [25, 26], the blue dashed lines bind the area between the second and the first 3-quantiles, while the purple line the area between between the 975th and 25th permilles. For the 4 distributions, real values easily fit in the 95%; anyway, for NODF values the real values lie just outside the central third of the probability. Notice the similar distributions for $NODF_t$ and $NODF_p$, as explained in Eq S6 in Supporting Information in S1 File. In **d)** the distribution for the assortativity values (obtained employing the definition by [33]): Even if the distribution is quite weird, the value measured on the real matrix is just outside the area containing the 33% of the distribution.

doi:10.1371/journal.pone.0140420.g004

Motifs. Let us represent an entry in the M−matrix as a square which is empty ($\square$) if $m_{cp} = 0$,

while it is filled ($\blacksquare$) for $m_{cp} = 1$. The checkerboard score [34], i.e. the number of pattern $\blacksquare\square \atop \square\blacksquare$ and $\square\blacksquare \atop \blacksquare\square$ inside the M−matrix counts the mutually exclusive exported products for two different countries. As shown in Fig 5(a), the total number of checkerboards is finely reproduced by the model; the agreement with real data means that the evolutionary algorithm well replicates the diversity of technological development roads followed by countries.

In [35, 36] several motifs for bipartite networks have been proposed in order to uncover the structural properties of the system at hand; in the following we will consider just the simplest ones, V−motifs and Λ−motifs. In few words, they represent the number of co-occurrence of products in the basket of two different countries (the V−motifs) or the co-occurrence of countries in the set of the producers of two different products (Λ−motifs);
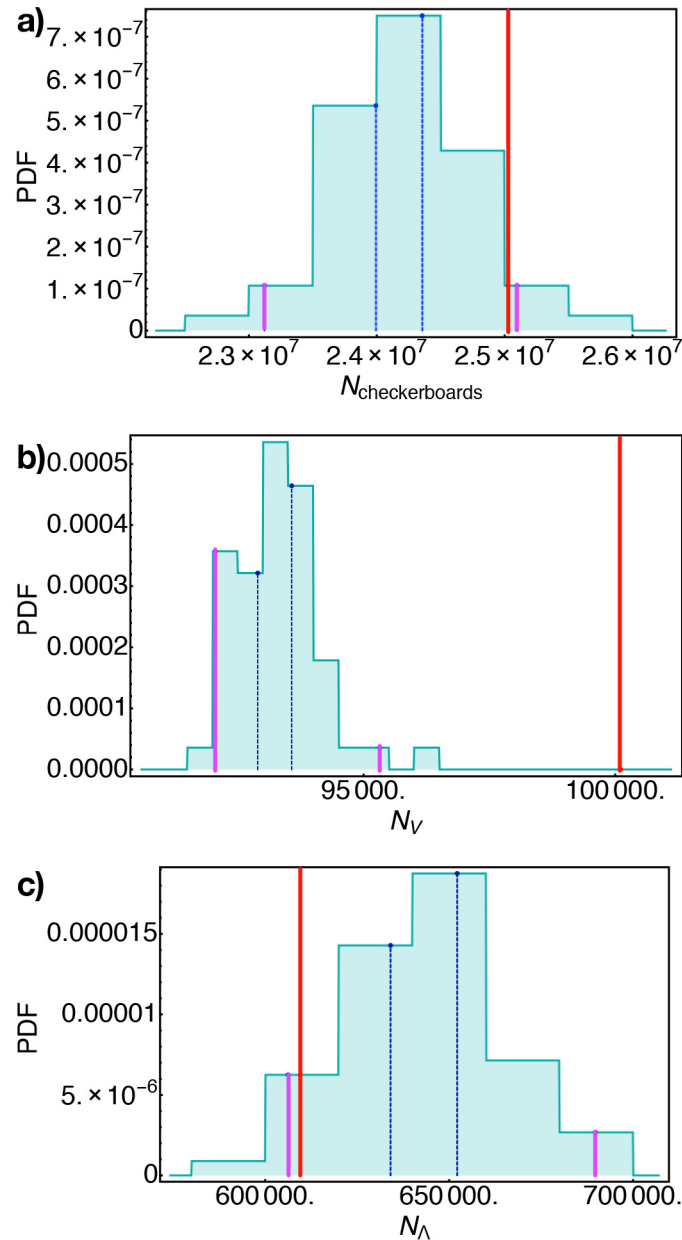
Fig 5. Checkerboards & Motifs. The distribution for the number of checkerboards(**(a)**), $V-$ (**(b)**) and $\Lambda-$ motifs (**(c)**), obtained from the simulation with initial conditions $N_{roots} = 20$ and $P_0 = 0.3$ and parameters $\alpha = 1.55$, $\beta = 0.8$, $\gamma = 0.3$, $k^0 = 4$. The red line is the observed value for the year 1980 from the dataset of [25, 26], the blue dashed lines bind the area between the second and the first 3-quantiles, while the purple line the area between between the 975th and 25th permilles. While **(a)** and **(c)** shows that the number of checkerboards $\Lambda-$motifs are reproduced by the model, in **(b)** the real value lies outside the 95% of probability; the presence of a hierarchy in the set of products captures the right values of checkerboards and $\Lambda-$motifs, but it is not enough to reproduce the $V-$motifs.

more details on the definitions of the motifs can be found in the Supplementary information.

In most of the cases the real value of the number of $\Lambda-$motifs falls at the borders of the first 3-quantile and far inside the area contained between the 25th and the 975th permilles (that
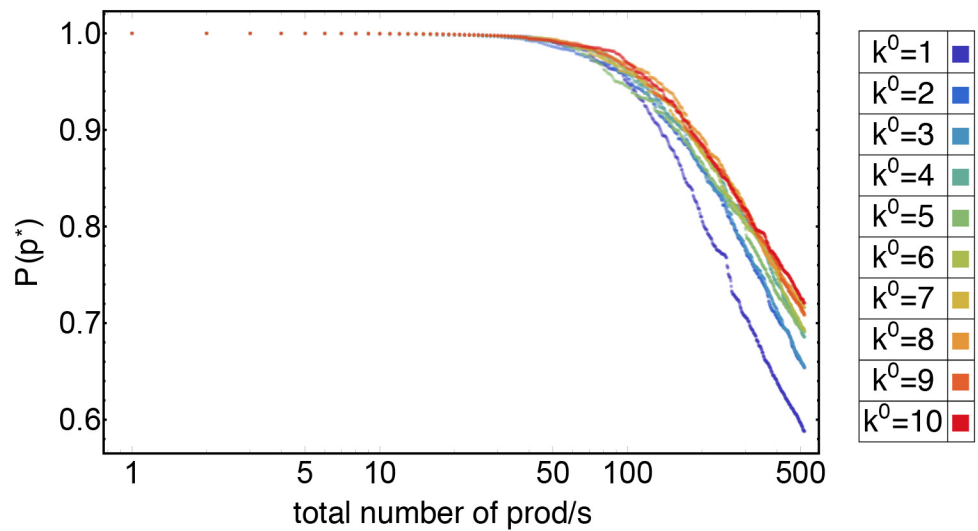
Fig 6. Probability of innovation. The evolution of the probability of innovation depending on k^0 (on the horizontal axis the total number of products at the time). It is possible to observe two different phases: a first period of discoveries, when the probability of innovating is close to 1 and a second period in which the spread of the novelties increase. It is worth noticing that all slopes for $k^0 > 1$ cluster together.

doi:10.1371/journal.pone.0140420.g006

is, the area containing the 95% of the probability around the mean value), see Fig 5(c). On the countries set we did not assume any kind of structure and it is probably the cause for not reproducing the total number of V−motifs, see Fig 5(b). In effect, simulated data always fall out of the 95% of the probability around the median of real data: apparently the evolutionary paths described by the products network is not enough for reproducing V−motifs.

The first 3 measurements, i.e i) the fitness and complexity scattered against the nodes degrees, ii) the nestedness and iii) the assortativity, explicit the hidden information encoded in the triangular shape of the matrix $\hat{M}$, while the motifs carry part of the topological information of the bipartite network. Our evolutionary model is able to replicate those measures, showing that the products network created by our algorithm can be a good starting point to better understand the hidden forces which produce the main characteristics of the export bipartite network structure.

The evolution

The total probability of introducing an innovation can be obtained assembling Eqs (3), (4) and (5):

$$
\begin{aligned}
P(p^*) &= \sum_{c,p} P_1(c) P_2(c|p) P_3(p^*|c,p) \\
&= \sum_c \frac{k_c^\gamma}{\sum_{c'} k_{c'}^\alpha} \cdot \sum_p \frac{k_p^\beta}{\sum_{p'} m_{cp'} k_{p'}^\beta} \cdot \frac{(k^0)^\gamma}{(k^0)^\gamma + \sum_{p''} A_{pp''} (k^0 + k_{p''})^\gamma}
\end{aligned}
\tag{6}
$$

The evolution of the probability of innovating respect to the total number of products exported at that time has been reported in Fig 6; different colors represent different values of k^0, while the other parameters (α, β, γ) have been fixed respectively to (1., 1.6, 0.4). At the early stage of
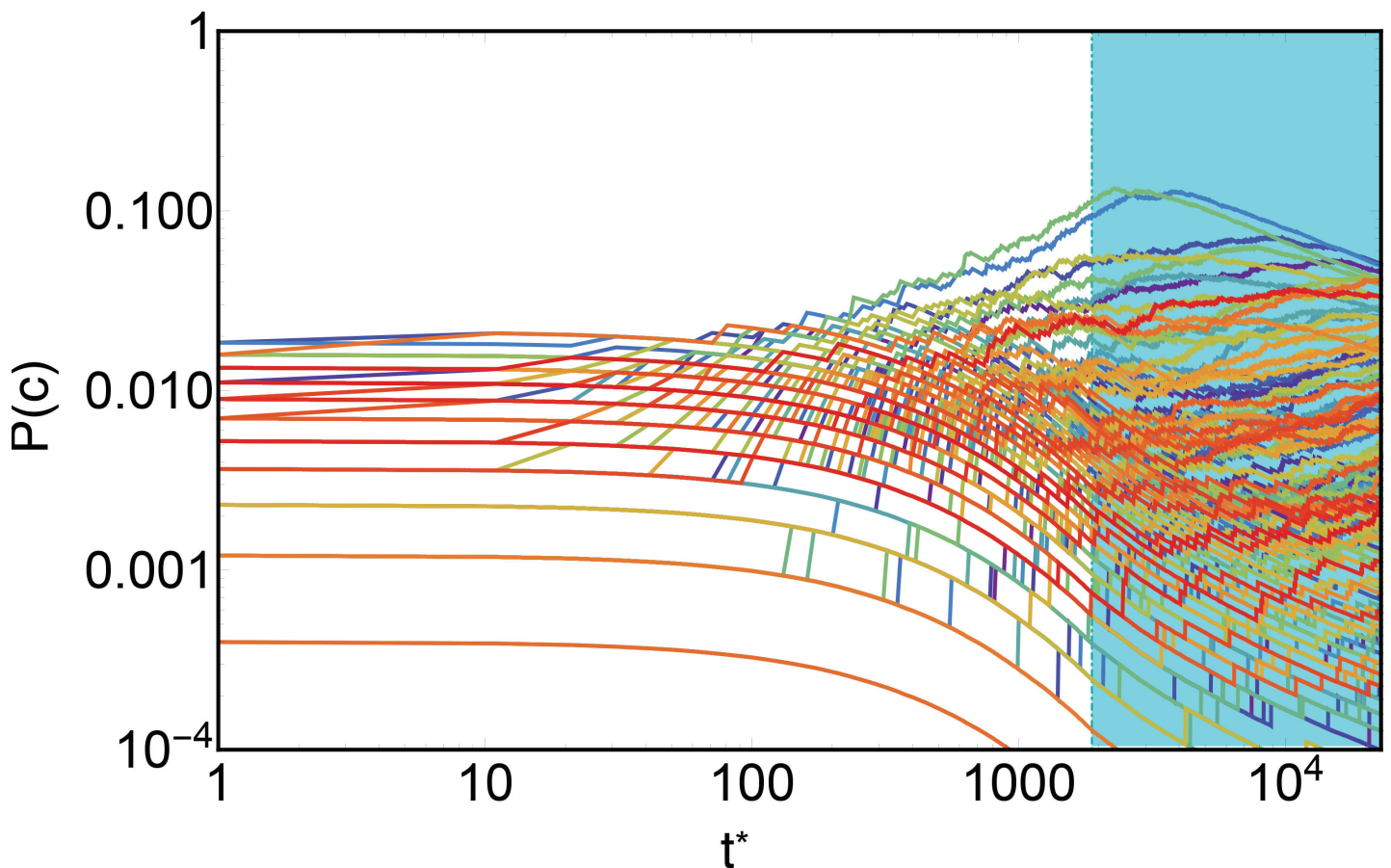
Fig 7. The evolution of the probability of selecting every country. On the horizontal axis there is the simulation time. Until the saturation regime (the cyan area) few countries start increasing their probabilities of being selected with the increasing of their diversification, to the detriment of the poor diversified countries, whose probabilities are pushed lower. In the saturation period, the mid-diversified countries enlarge their export basket, boosting their probabilities of being selected, while highest diversified countries are restrained; in this way the gap among countries reduces.

the evolution dynamics, the probability of innovating is obviously close to 1, as the number of branching is negligible with respect to the number of roots. This happens until the total number of products is around ~ 50 for all the value examined of k^0. After that threshold, the possibility of following path already developed by others countries reaches a higher value.

The plot in Fig 6 shows that there are two phases: a first period of "great discoveries" (until the total number of products is ≤ 50) in which the topology of the early products network is shaped, and a second period in which the technology innovations diffuse, under the form of novelties, among countries. Note also that slopes for different values of k^0 cluster: while $k^0 = 1$ is almost alone, sketching a steep trajectory, other values are next to each other.

A similar, but different, discussion can be made about the evolution of the probability of countries selection, given by Eq (3). In Fig 7 the evolution of this probability for single country is plotted (differently from Fig 6 on the horizontal axis the evolution time is plotted). This plot clearly shows the effect of the late saturation time interval (when, as explained above, $k^0 = 0$) on the selection probability of a single country: from an initial almost uniform diversification,

few countries start becoming more and more diversified, increasing in this way their probability of being selected at further times, and, consequently, reducing the same probability of poor ones due the normalization over all countries. In the saturation time interval, represented by the cyan area in Fig 7, we note a shrinking of the selection probability distribution over countries, due to the prevention of further innovation. Since just novelties are permitted at this final stage mid-diversified countries improve their chances of diversifying, while already developed are restrained by the fact that they are already extremely diversified and novelties occur with less frequency.

The same effect can be observed directly on the distribution of diversification over countries; Fig 8 shows that there is an abrupt raise in the diversification during the evolution for few countries, while others experience a slower evolution (again, on the horizontal axis there are the time steps of the simulation). Moreover, the most diversified countries, i.e. those which feel strongly the decrease in the probability of being selected due to the saturation, show an S–shaped profile of the evolution curve, Fig 8(b): after a steep raising slope, the saturation time determines a slow increase towards a limiting value. Fig 8(c) illustrates that not all countries show such a S–shaped behaviour due to the presence of the saturation stage, but just those with high diversification; others do not occupy the products network enough for feeling the difference between the products network growth and the saturation regime. The presence of such an S–shape curve is also typical in evolutionary models in biology and socio-economics, when resources are limited, [40, 41].

Discussion

The main target of the Economic Complexity approach, [11–19], is to unveil, through the information contained in the binary bipartite network of countries and exported products, the productive capabilities of different countries and the industrial hierarchical space. Quite surprisingly with respect to some celebrated economic theories, the biadjacency matrix $\hat{M}$, defining the bipartite network, exhibits a peculiar approximately triangular shape. This shows that the most diversified countries, i.e. those able to export a wide class of different products, are the only ones able to export both the most technologically advanced goods as well as the simplest ones. On the contrary poorly diversified countries usually export only ubiquitous products, which in general bring a low level of industrial complexity. This new approach permitted to reach many interesting results about the competitiveness of countries and the complexity of products. In particular the construction of a products network, defining a hierarchy among products, permits to determine the different paths followed by countries in the product space to develop their export basket [12, 18].

In this framework we proposed a simple dynamic evolution algorithm that, starting from general initial conditions is able to reproduce the main features of the observed countries/products bipartite network, as different measures for the "triangular" shape of the observed M–matrix, in a wide range of the parameters around reasonable values. The central ingredient of the evolutionary model is the progressive and self consistent construction of the product network, encoding the different steps of the technological progress.

Our model provides the simultaneous evolution of the matrix $\hat{M}$ and the products network; the dynamical evolution of the latter at the same time drives and is driven by the evolution of the matrix $\hat{M}$, as the technological evolution depends tightly on the productive capabilities of the different countries, i.e. on the nodes each country occupies in the products network.

The proposed model is able to reproduce the main features of the observed bipartite network for a wide range of parameters. In particular we compared the "shape" of the matrix $\hat{M}$, as encoded in the Fitness/Complexity algorithm [14–17, 19]; the measures of the nestedness
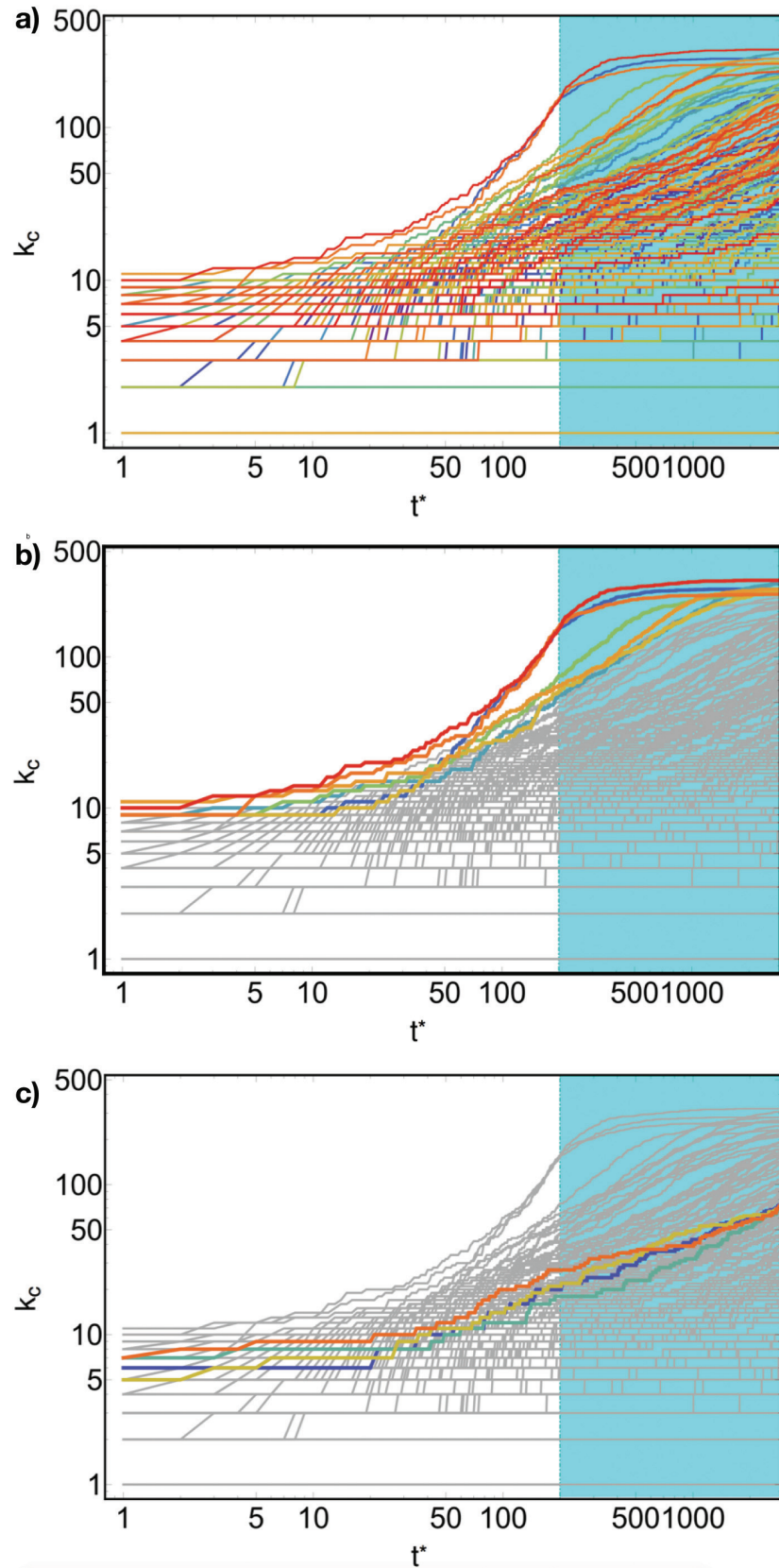
Fig 8. The evolution of the diversification against the simulation time. In **(a)** all countries, but those whose initial conditions have been particularly unlucky, evolve, boosting their diversification; in the cyan area, i.e. during the saturation regime, highly diversified countries experience an evolution different from others. In effect, it is possible to observe, once focusing on the highest fitness countries, **(b)**, that all of them experience a growth with a *S*–curve profile, which is peculiar of biologic and economics system with finite resources. Note the other less diversified countries, **(c)**, this phenomenon is not present; effectively the main target of the saturation regime is reducing the gap between the diversification of fully developed and developing countries.

doi:10.1371/journal.pone.0140420.g008

(in the definition proposed [32]) and the assortativity, [33]; some motifs, like the checker-boards patterns, [34] and the *V* and Λ, [35].

For the range of parameters examined we find the observed values of all quantities with a single minor exception (*V*–motifs) inside the 95% of the distribution of simulated data.

Our model is meant to be a first step in the direction of a dynamical network approach to the processes of countries innovation and competition on the exports. There are several possible directions of improvement, as implementing different evolutionary rules for the construction of the product network and the modeling of the countries dynamics on it.

Another possible direction could be the introduction of an appropriate random process of losing products from the export basket, simulating exogenous phenomenon as the progressive obsolescence of "old" products or the presence of socio-political factors (as wars, traditions, political resolutions, etc.). All these (and other) possible approaches are going to be studied in following works.

Supporting Information

S1 File. Supporting Information File. The supporting information with the mathematical appendix, and the description on the tuning parameter methodology.
(PDF)

Acknowledgments

We acknowledge Tiziano Squartini, Andrea Tacchella for many interesting and important discussions. We are also grateful to Claudia Testi for useful cross-checks.

Author Contributions

Conceived and designed the experiments: RDC FS. Performed the experiments: RDC FS. Analyzed the data: FS RDC. Wrote the paper: FS RDC AG. Revised the manuscript: RDC AG LP FS.

References

1. Verona G & Ravasi D Unbundling dynamic capabilities: an exploratory study of continuous product innovation. Ind Corp Change. 2003; 12: 577–606 doi: 10.1093/icc/12.3.577

2. Henisz WJ. The institutional environment for economic growth. Economics & Politics. 2000; 12: 1–31. doi: 10.1111/1468-0343.00066

3. Ostrom E, Schroeder L, Wynne S. Institutional incentives and sustainable development: infrastructure policies in perspective. Westview Press; 1993.

4. Belmonte A, Di Clemente R & Buldyrev SV. The italian primary school-size distribution and the city-size: a complex nexus. Sci Rep. 2014; 4. doi: 10.1038/srep05301 PMID: 24954714

5. Murtha TP & Lenway SA. Country capabilities and the strategic state: How national political institutions affect multinational corporations' strategies. Strateg Manage J. 1994; 15: 113–129. doi: 10.1002/smj.4250151008

6. Di Clemente R, Chiarotti G, Cristelli M, Tacchella A & Pietronero L. Diversification versus specialization in complex ecosystems. PloS one. 2014; 9: e112525. doi: 10.1371/journal.pone.0112525 PMID: 25384059

7. Maddison A. The world economy volume 1: A millennial perspective volume 2: Historical statistics Academic Foundation, 2007.

8. Pritchett L. Divergence, big time. J Econ Perspect. 1997: 3–17.

9. Aghion P, Howitt P & García-Peñalosa C. Endogenous growth theory MIT press, 1998.

10. Barro RJ & Sala-i Martin X. Economic growth, 2nd (2003).

11. Hausmann R & Klinger B. The structure of the product space and the evolution of comparative advantage. Center for International Development at Harvard University, 2007.

12. Hidalgo CA, Klinger B, Barabási AL & Hausmann R. The product space conditions the development of nations. Science. 2007; 317: 482–487. doi: 10.1126/science.1144581 PMID: 17656717

13. Hidalgo CA & Hausmann R. The building blocks of economic complexity. P Natl Acad Sci Usa. 2009; 106: 10570–10575. doi: 10.1073/pnas.0900943106

14. Caldarelli G et al. A Network Analysis of Countries' Export Flows: Firm Grounds for the Building Blocks of the Economy. PLoS ONE. 2012; 7: 47278. doi: 10.1371/journal.pone.0047278

15. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A & Pietronero L. A new metrics for countries' fitness and products' complexity. Sci Rep. 2012; 2. doi: 10.1038/srep00723 PMID: 23056915

16. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A & Pietronero L. Economic complexity: conceptual grounding of a new metrics for global competitiveness. J Econ Dyn Control. 2013; 37: 1683–1691. doi: 10.1016/j.jedc.2013.04.006

17. Cristelli M, Gabrielli A, Tacchella A, Caldarelli G & Pietronero L. Measuring the intangibles: A metrics for the economic complexity of countries and products. PloS one. 2013; 8: e70726. doi: 10.1371/journal.pone.0070726 PMID: 23940633

18. Zaccaria A, Cristelli M, Tacchella A & Pietronero L. How the taxonomy of products drives the economic development of countries. PloS one. 2014; 9: e113770. doi: 10.1371/journal.pone.0113770 PMID: 25486526

19. Pugliese E, Zaccaria A & Pietronero L. On the convergence of the Fitness-Complexity Algorithm. Preprint. Available: ArXiv:1410.0249. Accessed 06 October 2015.

20. Balassa B. 'revealed' comparative advantage revisited: An analysis of relative export shares of the industrial countries, 1953–1971. Manch Sch. 1977; 45: 327–344. doi: 10.1111/j.1467-9957.1977.tb00701.x

21. Albert R & Barabási AL. Statistical mechanics of complex networks. Rev Mod Phys. 2002; 74: 47–97. doi: 10.1103/RevModPhys.74.47

22. Newman M. The structure and function of complex networks. SIAM Rev. 2003; 45: 167–256. doi: 10.1137/S003614450342480

23. Caldarelli G. Scale-free networks: complex webs in nature and technology. OUP Catalogue, 2007.

24. Tria F, Loreto V, Servedio VDP & Strogatz SH. The dynamics of correlated novelties. Sci Rep. 2014; 4. doi: 10.1038/srep05890 PMID: 25080941

25. http://www.nber.org/data/ Accessed 06 Oct 2015.

26. Feenstra RC, Lipsey RE, Deng H, Ma AC & Mo H. World trade flows: 1962-2000 (2005). URL http://www.nber.org/papers/w11040 Accessed 06 Oct 2015.

27. Hausmann R & Hidalgo CA. HKS Working Paper No. RWP10-045, 2010.

28. Kauffman S. At home in the universe: The search for the laws of self-organization and complexity. Oxford University Press, 1995.

29. Krapivsky P & Redner S. Organization of growing random networks 2001.

30. Hesse H. Export diversification and economic growth. World Bank, Washington, DC, 2006.

31. Clauset A, Shalizi CR & Newman ME Power-law distributions in empirical data. SIAM Rev. 2009; 51: 661–703. doi: 10.1137/070710111

32. Almeida-Neto M, Guimaraes P, PRG Loyola RD Jr & Ulrich W. A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. Oikos. 2008; 117: 1227–1239. doi: 10.1111/j.0030-1299.2008.16644.x

33. Newman M. Assortative mixing in networks. Phys Rev Lett. 2002; 89: 208701. doi: 10.1103/PhysRevLett.89.208701 PMID: 12443515

34. Stone L & Roberts A. The checkerboard score and species distributions. Oecologia. 1990; 85: 74–79. doi: 10.1007/BF00317345

35. Saracco F, Di Clemente R, Gabrielli A & Squartini T Randomizing bipartite networks: the case of the world trade web. Sci Rep. 2015; 5. doi: 10.1038/srep10595 PMID: 26029820

36. Saracco F, Di Clemente R, Gabrielli A & Squartini T Detecting the bipartite World Trade Web evolution across 2007: a motifs-based analysis. Preprint. Available: ArXiv:1508.03533, Accessed 06 October 2015.

37. Bastolla U, et al. The architecture of mutualistic networks minimizes competition and increases biodiversity. Nature. 2009; 458: 1018–1020. doi: 10.1038/nature07950 PMID: 19396144

38. Suweis S, Simini F, Banavar JR & Maritan A. Emergence of structural and dynamical properties of ecological mutualistic networks. Nature. 2013; 500: 449–452. doi: 10.1038/nature12438 PMID: 23969462

39. Johnson S, Dominguez-Garcia V & Munoz MA. Factors determining nestedness in complex networks. Preprint. Available: ArXiv:1307.4685 Accessed 06 October, 2015.

40. Rogers EM. Diffusion of innovations. Simon and Schuster, 2010.

41. Barbosa LA & Fontanari JF. Culture–area relation in axelrod's model for culture dissemination. Theor Biosci. 2009; 128: 205–210. doi: 10.1007/s12064-009-0066-z