




Article

# Reconstructing Nonparametric Productivity Networks

Moriah B. Bostian <sup>1</sup>, Cinzia Daraio <sup>2,\*</sup> , Rolf Färe <sup>3,4</sup>, Shawna Grosskopf <sup>4</sup> , Maria Grazia Izzo <sup>2,5</sup>, Luca Leuzzi <sup>6,7</sup>, Giancarlo Ruocco <sup>5,7</sup>  and William L. Weber <sup>8</sup>

<sup>1</sup> Department of Economics, Lewis and Clark College, Portland, OR 97219, USA; mbbostian@lclark.edu

<sup>2</sup> Department of Computer, Control and Management Engineering A. Ruberti (DIAG), Sapienza University of Rome, 00185 Rome, Italy; mariagraziaizzo@gmail.com

<sup>3</sup> Department of Applied Economics, Oregon State University, Corvallis, OR 97331, USA; rolf.fare@oregonstate.edu

<sup>4</sup> Department of Economics, Oregon State University, Corvallis, OR 97331, USA; shawna.grosskopf@oregonstate.edu

<sup>5</sup> Center for Life Nano Science, Fondazione Istituto Italiano di Tecnologia (IIT), 16163 Rome, Italy; Giancarlo.Ruocco@roma1.infn.it

<sup>6</sup> Soft and Living Matter Lab, Institute of Nanotechnology, 00161 Rome, Italy; luca.leuzzi@roma1.infn.it

<sup>7</sup> Department of Physics, Sapienza University of Rome, 00185 Rome, Italy

<sup>8</sup> Department of Accounting, Economics and Finance, Southeast Missouri State University, Cape Girardeau, MO 63701, USA; wlweber@semo.edu

\* Correspondence: daraio@diag.uniroma1.it; Tel.: +39-06-77274-068; Fax: +39-06-77274-074

Received: 2 November 2020; Accepted: 7 December 2020; Published: 11 December 2020



**Abstract:** Network models provide a general representation of inter-connected system dynamics. This ability to connect systems has led to a proliferation of network models for economic productivity analysis, primarily estimated non-parametrically using Data Envelopment Analysis (DEA). While network DEA models can be used to measure system performance, they lack a statistical framework for inference, due in part to the complex structure of network processes. We fill this gap by developing a general framework to infer the network structure in a Bayesian sense, in order to better understand the underlying relationships driving system performance. Our approach draws on recent advances in information science, machine learning and statistical inference from the physics of complex systems to estimate unobserved network linkages. To illustrate, we apply our framework to analyze the production of knowledge, via own and cross-disciplinary research, for a world-country panel of bibliometric data. We find significant interactions between related disciplinary research output, both in terms of quantity and quality. In the context of research productivity, our results on cross-disciplinary linkages could be used to better target research funding across disciplines and institutions. More generally, our framework for inferring the underlying network production technology could be applied to both public and private settings which entail spillovers, including intra- and inter-firm managerial decisions and public agency coordination. This framework also provides a systematic approach to model selection when the underlying network structure is unknown.

**Keywords:** networks; data envelopment analysis; entropy; generalized multicomponent Ising model; Georgesçu-Roegen flows and funds model; Bayesian statistics; knowledge production

## 1. Introduction

Economic production often results from complex systems of inter-connected production processes, forming a unified network production technology. Data Envelopment Analysis (DEA) methods have long been used to estimate production technologies and measure relative performance. Network DEA (NDEA) models provide a generalization to assess the performance of complex systems, in which

separate production subtechnologies are linked. Network linkages include intermediate output/input relationships [1–4], as well as dynamic connections across time [5,6] and location [7–9], and two-way feedback effects between subtechnologies [10]. Chen et al. [11] consider potential pitfalls related to subtechnology efficiency, frontier projection, and the treatment of intermediate quantities when using envelopment methods. Chen et al. [12] and Cook et al. [13] review the literature for two-stage network models. Kao [14] provides a more recent, comprehensive review of the NDEA literature. Several handbooks [13,15] cover topics related to practical implementation of NDEA.

While NDEA models can accommodate a large number of sub-activities and interactions, the curse of dimensionality requires that the number of sub-activities/interactions be relatively small compared to the number of observations for NDEA to discriminate differences in performance. This poses a tradeoff between better structurally representing system dynamics and better assessment of relative performance. To avoid the curse of dimensionality, we estimate a relatively parsimonious NDEA model and then use the resulting performance estimates as inputs into an entropy-based statistical model that allows recovery of a wider set of network interactions.

Uncertainty surrounding model selection poses another common limitation for NDEA. Kao [16] argues that it is not possible to *a priori* choose which network model to apply in a given empirical context. Most existing studies analyze productivity networks in a descriptive way, without considering them in a statistical framework. As a result, the network structure is generally assumed and not estimated. Trinh and Zelenyuk [17] offer one exception, proposing a bootstrap-based comparison between average DEA-NDEA efficiency scores and their distributions, but without questioning the network structure, i.e., assuming the NDEA structure.

We propose a statistical-based approach to reconstruct (infer) the network's structure, for nonparametrically-estimated productivity frontier models. We develop a Bayesian framework that relies on recent Pseudo-Likelihood techniques introduced in the physics of complex systems [18–21] for estimation. Our approach can be considered semi-parametric because it bases inference of the network structure on a parametric Bayesian approach (generalized multicomponent spin model) to make inference for nonparametric NDEA productivity networks.

Up to now, statistical inference for NDEA models has been constrained by the lack of standard inferential tools needed to evaluate complex systems. We provide a reasonable and robust inferential approach that allows us to reconstruct productivity networks empirically, starting from the observed NDEA data. The expected new insight gained by applying our methodology is then to infer the productivity network structure from the data instead of assuming it *a priori*. Although our inferential approach is parametric and relies on various parametric assumptions and Bayesian statistics, its foundation is laid by Georgesçu-Roegen [22] in which the economic process follows the second law of thermodynamics—the entropy law.

Our work also closely relates to recent developments in the econometrics of information [23–25], statistical inference, and machine learning [26,27]. By including work from information science and machine learning methods borrowed from the physics of complex systems, we fill an existing gap in the NDEA literature related to the choice of the network structure, described as an open issue in recent books on the subject [15,16]. Ours is also a timely contribution, as the diffusion of computational power of computers permits the implementation of new inferential tools based on machine learning techniques.

Our empirical illustration examines cross-country and cross-discipline knowledge production spanning 16 STEM (Science, Technology, Engineering, Math) fields and 17 years. We estimate research linkages across individual fields, both within and across countries over time, to assess research performance in both quantitative and qualitative terms. Our key insight from the application is that while estimated efficiency measures for research output exhibit generally low correlations across disciplines, we find that many of these disciplines exhibit relatively high interdependencies. Simple correlation measures fail to capture underlying structural relationships connecting research disciplines.

Our framework for inference extends previous applications that model knowledge networks through NDEA (see [6,28–31]). Daraio [32] shows that the complexity of research productivity and

the expansion of networks in economics provides impetus for the search for new and more general models of the production process. Our application extends Daraio et al. [33] to estimate scientific knowledge productivity, offering a more general network model that accounts for the complexity of research production.

The paper unfolds as follows. In the next section, we illustrate the main features of the economic model. Section 3 presents the axioms of the underlying DEA models and their connection to those of the general NDEA model. Included is a schematic of the general types of structures of NDEA models. This section also introduces the Georgescu-Roegen [22] flow and funds model (GRFF) and its connection with our NDEA model. Section 4 illustrates the connection between the statistical approach proposed and the GRFF model. Section 5 introduces the knowledge production network that we estimate, including a schematic of the possible cross-disciplinary links that our statistical second stage estimates can reveal. Next is a description and summary statistics of the data, followed by the outlines of the alternative parsimonious NDEA models we estimate in the first stage. A formal statement of the NDEA problem objective and constraints follows. Descriptive analyses of the first stage productivity models are included, followed by the main results of the application of the second stage to our knowledge production. The final section provides a discussion of our approach and results. We include two technical appendices: Appendix A contains an introduction to the Ising spin glass model while Appendix B provides additional technical details on the Pseudo-likelihood approach. We also include a more detailed summary of our data in Appendix C.

## 2. The Economic Model

The axiomatic production theory behind this paper and described in Section 3, is found in [1,34] and [35]. Färe and Grosskopf [3] introduce the concept of NDEA and extend the axioms to a network setting. Section 3.2 highlights the correspondence of the axiomatics of NDEA with the representation of the production process with flows and funds a la Georgescu-Roegen. This correspondence yields a new, more general framework for modeling production processes by integrating the production process, information theoretic approaches to econometrics, machine learning and statistical inference from the physics of complex systems.

As described by Prieto and Zofio [4], NDEA within an input-output model (for an introduction and a deep overview see [36]) allows us to gauge potential productivity gains by comparing technologies corresponding to different “economies”. Such models represent a network where different sectoral nodes use primary inputs (endowments) to produce intermediate input and outputs (according to sectoral technologies). In graph theory terms, in an input-output model, each sector (industry) is represented by a node and each flow of intermediate inputs and outputs is represented by a link. Hence, it is possible to optimize primary input allocation, intermediate production and final production using NDEA. This framework allows us to model the different sub-technologies corresponding to alternative production processes, to assess efficient resource allocation among them, and determine potential output gains that could be realized by reducing inefficiencies. In this setting, we use statistical inference to estimate the chains/path connections within and between nodes in order to reveal the underlying structure of the input-output system (see [36], p. 675).

## 3. Axiomatics of DEA Network Models

### 3.1. DEA

#### 3.1.1. Basic Axioms

NDEA is widely used in economics and operations research to assess efficiency and productivity in complex technologies or systems. As of 23 November 2020 Google Scholar identifies 173,000 articles identified under ‘Network DEA’ between 1996 and 2020. It is an extension of Data Envelopment Analysis (DEA) which was introduced by Charnes, Cooper and Rhodes [37]. As its name suggests, DEA envelops input-output data to identify the best practice ‘frontier of technology’ in the sample data. Individual data points are compared to that best practice frontier to determine relative performance.

A well-known issue associated with DEA is the curse of dimensionality: adding more inputs and outputs to the model requires more data to discriminate performance. DEA as an estimator has a slow rate of convergence, making statistical inference difficult. This difficulty is inherited by and compounded in NDEA. NDEA has the additional issue of inference concerning the structure of the more complicated model structure, which connects multiple subtechnologies. This can be structured in many ways, i.e., there are many potential models. This motivates our research question: is there a statistical way to infer the structure of the network model? And can we use that as guidance in model selection when we are forced to specify fairly parsimonious network models due to the lack of data and curse of dimensionality. Our contribution is the derivation and application of just such a statistical approach to structure inference and model selection for NDEA.

NDEA models are a generalization of the basic DEA or activity analysis models of technology and efficiency; they are often referred to as looking inside the black box technology assumed in more reduced-form DEA efficiency models. NDEA can be used to model production processes when the choice of inputs/outputs in one period affects what can be produced in subsequent periods. In addition, NDEA can be used to model production processes where intermediate products are produced in one stage of production and are then used to produce final outputs in another stage. They are also useful when production entails spillovers that can enhance (as in our knowledge production application) or detract from production by other producers/DMUs.

Following [1], we show that the axiomatic underpinnings are similar to those of the standard DEA-estimated technology. Notationally, for inputs,  $x \in \mathbb{R}_+^G$ , and outputs,  $y \in \mathbb{R}_+^M$ , we define the graph of technology, or production set, which relates inputs to outputs:

$$GR = \{(x, y) : x \text{ can produce } y\}.$$

Depending on the problem at hand, we can model the technology equivalently in terms of the input set:

$$L(y) = \{x : (x, y) \in GR\}$$

or the output set,

$$P(x) = \{y : (x, y) \in GR\}.$$

For estimation, activity analysis (DEA) models generally employ a set of linear constraints on the inputs and outputs to construct the so-called piecewise linear frontier of the technology set, whether  $GR$ ,  $L(y)$  or  $P(x)$ , in accordance with the basic axioms of production theory (listed below and following [35]).

Following the terminology coined by Charnes, Cooper and Rhodes [37] let there be  $D$  activities or Decision Making Units (DMUs) i.e.,  $\gamma = 1, \dots, D$ . Each activity (DMU) has an associated input–output vector  $(x^\gamma, y^\gamma) = (x_{\gamma 1}, \dots, x_{\gamma G}, y_{\gamma 1}, \dots, y_{\gamma M})$ . Kemeny, Morgenstern and Thompson [38] relaxed the von Neumann [39] axioms that all inputs/outputs for each DMU be strictly positive and proposed the following non-negativity conditions on the input and output data, where the data are sometimes referred to as ‘coefficients’. These conditions essentially require that the data matrix be of full rank. These include:

- $\sum_{m=1}^M y_{\gamma m} > 0$ ,  $\gamma = 1, \dots, D$ , each DMU produces some output;
- $\sum_{\gamma=1}^D y_{\gamma m} > 0$ ,  $m = 1, \dots, M$ , each output is produced by some DMU;
- $\sum_{\gamma=1}^D x_{\gamma g} > 0$ ,  $g = 1, \dots, G$  each input is used by some DMU;
- $\sum_{g=1}^G x_{\gamma g} > 0$ ,  $\gamma = 1, \dots, D$  each DMU uses some input.

If these assumptions are satisfied, then following [35], Färe and Grosskopf [1] show that the basic activity analysis (DEA) technology, here specified as an output set:

$$P(x) = \{y : y_m \leq \sum_{\gamma=1}^D \lambda_{\gamma} y_{\gamma m}, \quad m = 1, \dots, M, \\ \sum_{\gamma=1}^D \lambda_{\gamma} x_{\gamma g} \leq x_g, \quad g = 1, \dots, G, \\ \lambda_{\gamma} \geq 0, \quad \gamma = 1, \dots, D\} \quad (1)$$

satisfies the axiom set below, which provides a minimal set consistent with neoclassical production theory. We note that the  $\lambda$  variables are so-called intensity variables which serve to ‘construct’ the piecewise linear frontier of the technology/output set. A typical DEA application using the constraints in  $P(x)$  above (1), seeks to maximize outputs for each DMU, subject to the input and output constraints based on the entire data set. This yields an efficiency score for each DMU, where a value of unity signals best practice performance.

The basic production axioms include:

- A.1  $0 \in P(x)$ ,  $\forall x \in \mathbb{R}_+^G$ ,  $y \notin P(0)$ ,  $y \geq 0$ . (inactivity is feasible);
- A.2  $x \in L(y)$ ,  $\lambda \geq 1 \Rightarrow \lambda x \in L(y)$  (weak disposability of input);
- A.2S  $x \geq x^o \in L(y) \Rightarrow x \in L(y)$  (strong disposability of input);
- A.3  $y \in P(x)$ ,  $0 \leq \theta \leq 1 \Rightarrow \theta y \in P(x)$ ;
- A.3S  $y \leq y^o \in P(x) \Rightarrow y \in P(x)$  (strong disposability of output);
- A.4  $\forall x \in \mathbb{R}_+^G$ ,  $P(x)$  is bounded;
- A.5 The graph is a closed set.

These are minimal axioms consistent with neoclassical production theory. A.1 allows for inactivity, A.2–A3S describe feasible constraints on inputs and outputs, imposed through the respective inequalities imposed in the input and output constraints in  $P(x)$ . A.4 requires that DMUs cannot produce unlimited output with given inputs, and A.5 requires that the graph technology set contain its boundary, which then serves to identify best practice.

In addition, it is often convenient to assume convexity of the input and output sets. The general result is that if each subtechnology in the network satisfies the Kemeny et al. conditions [38], then the network satisfies the axioms above. Similarly, if each subtechnology exhibits constant returns to scale, then the network also exhibits constant returns. We note that this holds for directed networks.

### 3.1.2. What Makes a Network?

We introduce the general structure of the network model with a figure first introduced by [40], which illustrates several types of networks. See also [41]. The box in the figure represents the basic DEA models with exogenous inputs  $x_o$  entering the ‘black box’ producing outputs  $y^4$  exiting the black box technology. Ignoring the interior of the box would be consistent with the DEA technology described by the linear constraints in (1). The network model allows specification of multiple processes or subtechnologies. Here we assume that there are three sub-technologies

$$P^1, P^2, P^3$$

organized as in Figure 1, where outputs from  $P^1$  and  $P^2$  enter  $P^3$  as inputs.

We extend this model to include a source, ‘o’, which distributes inputs to the network and a sink, ‘4’, which collects the network outputs. The notation identifies the source of the variable with a subscript and the destination with a superscript. So  $x_o^i$ ,  $i = 1, 2, 3$  means that input  $x_o$  is distributed to the three subtechnologies, and we have

$$x_o \geq x_o^1 + x_o^2 + x_o^3.$$

Similar notation is used for outputs, so  $y_2^3$  means that outputs from  $P^2$  are inputs into  $P^3$ . The final output is

$$y = y_1^4 + y_2^4 + y_3^4,$$

with the appropriate choice of dimension of the output vectors. This schematic includes the possibility of parallel subtechnologies or processes such as  $P^1$  and  $P^2$ , as well as sequential sub-technologies which could be linked through time providing a basis for a dynamic network or supply chain, echoing the earlier work of Georgesçu-Roegen.

A formal mathematical statement of the network problem we solve in our illustration is deferred to Section 5.

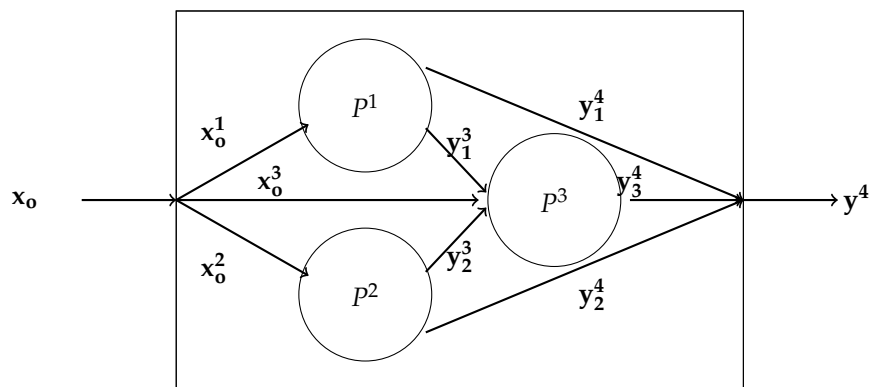


Figure 1. Three node network with input source and output sink.

### 3.2. Connection with Georgesçu-Roegen’s Flows and Funds Model

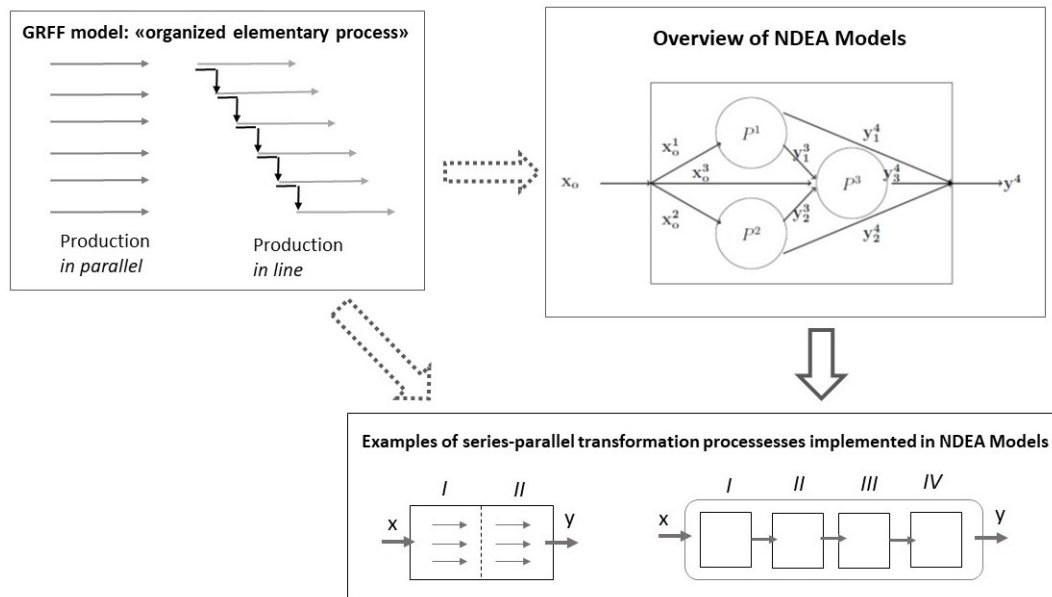
The network model analyses the joint actions of different activities within a process. Our theoretical framework allows us to analyze and represent production processes much such as the Georgesçu-Roegen Flows and Funds model (hereafter GRFF model) as subtechnologies which are connected to form the broader network via a maximum entropy condition. In this section, we show how the GRFF Model bridges the axiomatic of NDEA and estimation techniques based on complex systems.

NDEA models use the structure of networks to model production processes. Georgesçu-Roegen in the 1970s proposed a production model based on “organized elementary process” which can be in line or in parallel. We observe here that this production element is implicitly used in NDEA models. The “organized elementary process” of the GRFF model is the main ingredient or kernel of the axiomatics of NDEA introduced in Section 3 and of the transformation processes modelled in the NDEA literature. We think we are the first to point out the correspondence illustrated in Figure 2.

Figure 2 contains three panels. The North-West panel shows the elementary unit of the Georgesçu-Roegen production model, the so-called “organized elementary process” which can be of parallel or in-line production. The North-East panel illustrates the Network DEA models that we presented in the previous section, characterized by both parallel ( $P^1$  and  $P^2$ ) and in-line production processes (such as  $P^1$  and  $P^3$  or  $P^2$  and  $P^3$ ). The South panel shows two examples of processes modeled in NDEA models. The model on the left illustrates a two-stage production process in parallel while the model on the right shows a four-stage production process in line. As mentioned above, GRFF’s “organized elementary process” is implied in the North-East and South panels and for this reason, in Figure 2, we reported in dashed form the arrows from the North-West panel towards the other two panels. In this way, we highlight how GRFF’s “organized elementary process” is implicitly contained in the other two panels.

The schematic representation of NDEA and the possibility of including both parallel and sequential sub-technologies can be linked then to the GRFF model. Generally (see e.g., [42]), the model takes into account the actual characteristics of production elements and processes, such as, indivisibility, complementarity, tacitness and heterogeneity of productive knowledge.

In the GRFF model, a flow is an input or an output that enters or exits from a process (for example, energy, water, software, loom, computer, etc.). A fund provides its services to several processes that occur over time (for example, worker, software, land, loom, computer, etc.). A distinction is made between the agents of production processes and the services that they provide. Activities consist of different operations which require the performance of one or more elementary tasks. An elementary task is an operation which, by definition, is not further divisible (for instance, loading or unloading an intermediate product or cutting a piece of fabric). The GRFF model can be implemented both at the microeconomic level, considering individual case studies, and at the macroeconomic level, analysing a set of production units in different sectors of activity.



**Figure 2.** Connections among GRFF Model, our Axiomatics of NDEA and examples of NDEA models .

The GRFF model allows the analytical representation of the organization of production processes including the organization and time dimension of production processes. The formulation of the network production technology presented in Section 3 is an implementation of the GRFF model. The GRFF model may also be connected to the neo-Schumpeterian interpretative framework of production of new processes by means of creation and diffusion of knowledge [43], in which there is an interplay between capabilities, transactions and scale and scope to explain the boundary and the competitiveness of the analyzed units [44].

#### 4. The Statistical Model

##### 4.1. Maximum Entropy and Georgesçu-Roegen

The principle of maximum entropy serves as the foundation of the theory of inference [45], providing the statistical mechanics to reconstruct probabilistic information from incomplete data. Physical systems evolve spontaneously and possess stability characteristics at equilibrium, which is characterized by the value of maximum entropy. The key to the application of the principle (see [45]) is associating to a probability density function (pdf) an entropy function that measures the dispersion or uncertainty with which the occurrence of possible events are expected. This allows us to introduce constraints, based on our knowledge of the system, that can be treated with the formalism of Lagrange multipliers (see Section 4.2 for more details).

Generally, entropy may be interpreted as: (i) a measure of disorder in a system, (ii) a measure of our ignorance of a system, and (iii) an indicator of the irreversible changes in a system [46].

The Austrian school posits the economy as a complex system that is the outcome of uncoordinated individual behavior [47,48]. In such systems, equilibria do not always refer to a “stationary state” but instead are related to the concept of attractors. An attractor is a deterministic sequence of states which are cyclically visited by the system. As such, it becomes impossible to fully understand macro processes by examining individual behavior. Although representatives of the Austrian School had a skeptical regard to the use of mathematical tools in economics [49], their ideas can be expressed through the lense of statistical thermodynamics or the theory of information (see also [50]).

In chapter VI of his 1971 book [22], Georgesçu-Roegen describes the introduction of statistical mechanics and highlights the connection of economic processes with the second law of thermodynamics, i.e., the entropy law.

We draw the connections here between statistics, economic productivity, and physics of complex systems. As is well known, the correlation between two variables can be influenced by other confounding factors, and does not imply a direct causal effect of one variable on another. On the other hand, interactions or interdependencies refer to strict relationships between variables, allowing us to describe the impact of the variation of one variable on another. Economic productivity models can be used to analyze these interdependencies in a production process, as well as the interconnections between different economic sectors. Physics investigates the interactions between particles in order to analyze direct and reciprocal effects.

We finally highlight the fact that in information theory the maximum entropy problem can be reformulated as a minimum cost of coding, which is actually a function defined as the opposite of the entropy [51].

We aim to derive the level and the structure of interactions between disciplinary research productivities. We can think of this as an inverse problem, as inference of the underlying network is drawn from observational data [25]. Importantly, Georgesçu-Roegen [22] provides the theoretical support for the unknown model parameters, and justification of the assumptions underlying our statistical model.

#### 4.2. Maximum Entropy Estimates

We define our variables as the vectors  $\mathbf{s}_i = (s_i^{(1)}, \dots, s_i^{(\gamma)}, \dots, s_i^{(D)})$ ,  $i = 1, \dots, N$ . Here and in the following bold style marks a vector quantity. Subscripts, e.g.,  $i$ , indicate disciplines, whereas the superscript index  $\gamma = 1, \dots, D$  refers to a given Country. The observable  $\mathbf{s}_i$  depends on the observation time  $t$ , thus the set of data  $\{\mathbf{s}(t)\} = \{\mathbf{s}_1(t), \dots, \mathbf{s}_i(t), \dots, \mathbf{s}_N(t)\}$  can be defined, where  $t = 1, \dots, T$  identifies a given realization of the generic set of variables or configuration  $\{\mathbf{s}\}$ . To simplify the notation, where unnecessary, the index  $t$  will be omitted. The element of our variable,  $s_i^{(\gamma)}$ , in the case of our application to knowledge production (see Section 5.1), is related to productivity in the discipline  $i$  of the Country  $\gamma$  as follows

$$s_i^{(\gamma)}(t) = \frac{\Delta_i^{(\gamma)}(t)}{\sqrt{\sum_{\gamma=1}^D \Delta_i^{(\gamma)}(t)^2}}; \Delta_i^{(\gamma)}(t) = \pi_i^{(\gamma)}(t) - \bar{\pi}_i(t); \tag{2}$$

$$\bar{\pi}_i(t) = \frac{1}{D} \sum_{\gamma=1}^D \pi_i^{(\gamma)}(t); \gamma = 1, \dots, D; i = 1, \dots, N; t = 1, \dots, T. \tag{3}$$

$\pi_i^{(\gamma)}(t)$  is the productivity of country  $\gamma$  in disciplinary subject category  $i$  at time  $t$ . We let  $\bar{\pi}_i(t)$  represent the world-country average of productivity in subject  $i$ , so that  $\bar{s}_i = 0$  and  $\bar{s}_i^2 = \frac{1}{N}$ . We can use this formulation to account for the recent trend of increasing worldwide scientific productivity, considering deviations from the world-country average productivity,  $\Delta_i^{(\gamma)}(t)$ , in place of  $\pi_i^{(\gamma)}(t)$ . While the average scientific productivity increases over time, the distribution of the deviations around the means does not. See the evidence reported in [52].



Shannon's [51] theorem states that entropy ( $S$ ), defined in statistical mechanics, is a measure of the 'amount of uncertainty' related to a given discrete probability distribution  $p(\mathbf{s})$ . Accordingly,  $S[p]$  is given by

$$S[p] = -K \sum_{\{\mathbf{s}\}} p(\mathbf{s}) \log[p(\{\mathbf{s}\})], \quad (4)$$

where  $K$  is a positive constant and  $p(\{\mathbf{s}\})$  is the pdf (probability density function) of the configuration  $\{\mathbf{s}\}$ . This quantity is positive, additive for independent sources of uncertainty and it agrees with the intuitive notion that a uniform (or broad) distribution represents more uncertainty than does a sharply peaked distribution. It is immediate to verify the latter observation in the one-dimensional case by considering Equation (4) and taking into account the property of the discrete distribution of probability,  $p_i \leq 1$ .

In making inference on the basis of incomplete information we must use that probability which maximizes the 'amount of uncertainty' or entropy subject to whatever is known [45]. This yields an unbiased assignment, avoiding arbitrary assumption of information which by hypothesis we do not have [45]. For a set of variables  $\{\mathbf{s}\}$  the so-called empirical expected value of a given function of  $\{\mathbf{s}\}$  is defined as the average of the function over the observed realization (the mean) of  $\{\mathbf{s}\}$ , in the present case the average over time. Since some empirical expectation values can be measured, formally this means that  $p(\{\mathbf{s}\})$  can be found as the solution of a constrained optimization problem, i.e., maximizing the entropy of the distribution subject to conditions that enforce the expected values to coincide with the empirical ones. We will refer to the quantities whose averages are constrained as 'features' of the system. For simplicity, we choose to observe the lower order statistics of the data which can bring information about the underlying network of interactions between variables, i.e., pairwise correlations. The features of the system we are considering are thus the two-variable combinations  $\mathbf{s}_i \cdot \mathbf{s}_j$ ,  $i, j = 1, \dots, N$ ,  $i \neq j$ . The optimization problem reduces to

$$\text{Max}_{p(\{\mathbf{s}\})} S[p], \quad (5)$$

with the constraints

$$\sum_{\{\mathbf{s}\}} p(\{\mathbf{s}\}) = 1 \text{ and } \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle = \frac{1}{T} \sum_{t=1}^T \mathbf{s}_i(t) \cdot \mathbf{s}_j(t), \quad i, j = 1, \dots, N, i \neq j, \quad (6)$$

where ' $\langle \rangle$ ' is the true average over the distribution  $p$ . The symbol ' $\cdot$ ' denotes the Hadamard product of two vectors. The first constraint accounts for the correct normalization of the pdf, whereas the second one arises from the required equality between true and empirical average of the above-defined features. Generally, if certain interdependencies are known to exist between the elements of the matrix  $S[p]$ , constraints can be imposed to account for these interdependencies. However, instead of imposing a priori interdependency constraints we chose to infer them instead of assuming their existence. Solving Equation (5) with the constraints (6) leads to

$$p(\{\mathbf{s}\}) = \frac{e^{-\frac{1}{2} \sum_{i \neq j} J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j}}{Z}. \quad (7)$$

The chosen parameters,  $J_{ij}$ , are symmetric, i.e.,  $J_{ij} = J_{ji}$  (There is a link between the assumption of equilibrium underlying a Boltzmann-Gibbs distribution, and symmetry of the pairwise interactions. Symmetric couplings lead to a steady state described by the Boltzmann-Gibbs distribution while asymmetric couplings lead to a non-equilibrium state [53]. We can assign to the system a particular dynamics, which leads it to a given steady state distribution. Recent developments achieved for dynamical inverse Ising model [54,55] could represent an interesting extension of the present work, which is left for future research). The constant  $Z$  can be determined by exploiting the constraint  $\sum_{\{\mathbf{s}\}} p(\{\mathbf{s}\}) = 1$ , obtaining  $Z = \sum_{\{\mathbf{s}\}} e^{-\sum_{i \neq j} J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j}$ . The parameters  $J_{ij}$  are determined by requiring the

second constraint to be fulfilled. Asymmetric  $J_{ij}$  can always be re-conducted to symmetric  $J_{ij}$ , which give rise to the same values of pdf if this latter is the Gibbs distribution in Equation (7).

We observe that the pdf defined in Equation (7) coincides with the Maxwell-Boltzmann probability distribution function at a given fixed temperature,

$$p(\{\mathbf{s}\}) = \frac{e^{-H(\{\mathbf{s}\})}}{Z} \quad (8)$$

related to an Ising model with spin  $\mathbf{s}_i$ , interaction parameters  $J_{ij}$  and zero magnetic field, described by the Hamiltonian

$$H = -\frac{1}{2} \sum_{i \neq j}^N J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j. \quad (9)$$

The quantity  $Z$  in Equations (7) and (8), constant with respect to  $\{\mathbf{s}\}$  but dependent on the set of parameters  $\{J\}$  with generic element  $J_{ij}$ , is called in this context the partition function.

The connection between maximum entropy and maximum likelihood is indeed well known (see e.g., [56]) and the pdf in Equation (7) which satisfies the constraint on the parameters  $J_{ij}$  given in Equation (6) can properly be derived by searching the maximum of the so-called Likelihood function within the class of models of the Boltzmann distribution related to an Ising model with zero magnetic field. The Likelihood function is defined in the context of Bayesian inference (see e.g., [26]). By assuming that (i) each realization of the set  $\{\mathbf{s}\}$  is drawn independently, (ii) the data have been generated by a (known) model, which depends on the set of (unknown) pairwise parameters  $\{J\}$ , one aims to find the optimal values of  $\{J\}$ , i.e., the ones which maximize the conditional probability  $p(\{J\}|\{\mathbf{s}\})$ . From the Bayes theorem [26] it follows that

$$p(\{J\}|\{\mathbf{s}\}) = \frac{p(\{\mathbf{s}\}|\{J\})p(\{J\})}{p(\{\mathbf{s}\})} = \frac{p(\{\mathbf{s}\}|\{J\})p(\{J\})}{\int_{\{J\}} p(\{\mathbf{s}\}|\{J\})p(\{J\})}. \quad (10)$$

The probability  $p(\{J\}|\{\mathbf{s}\})$  is called *posterior*,  $p(\{J\})$  *prior*,  $p(\{\mathbf{s}\})$  *evidence* and  $p(\{\mathbf{s}\}|\{J\})$  *Likelihood*. If the prior is the uniform distribution, as we assume here, the most probable a posteriori set of variables is, as a consequence of Equation (10), the one which maximizes the Likelihood function. Under the further assumption that the Likelihood function belongs to the class of Boltzmann distribution functions, the so-called Log-Likelihood function can be defined as

$$l(\{J\}) = \sum_{t=1}^T \log[p(\{\mathbf{s}(t)\}|\{J\})] = \sum_{t=1}^T -H(\{\mathbf{s}(t)\}|\{J\}) - T \log(Z(\{J\})). \quad (11)$$

If one assumes that the system can be described by an Ising-like, pairwise interacting model (see Appendix A for an introduction and additional details) with zero external field, the Hamiltonian  $H$  is the one defined in Equation (9). We observe that in the definition of the Log-Likelihood function in Equation (11) the hypothesis of independency of the realizations of the configuration  $\{\mathbf{s}\}$  at different times has been exploited, see point (i) above. Thus, the optimization problem reduces to choosing the set of parameters  $\{J\}$ , which maximize the pdf in Equation (11). A quick calculation of the first and second derivatives of Equation (11) with respect to the parameters  $J_{ij}$  shows that the set of parameters which maximizes Equation (11) should indeed maximize Equation (5) with the constraint in Equation (6).

The Ising model has been widely applied in different fields, such as modelling the behaviour of magnets in statistical physics [57], image processing and spatial statistics [58–60], modelling of neural networks [61] and social networks [62]. It is, however, worth noting that by exploiting Shannon's theorem the Ising model does not arise from specific hypotheses about the underlying network but instead is the least-structured model consistent with the measured pairwise correlations. In Appendix A, we outline how the Ising spin glass model is introduced in the physics of complex systems. Table 1 describes the main components of our model and the correspondence between the Ising spin model from statistical physics and economic productivity analysis.

**Table 1.** Model's components.

Statistical Physics	Productivity Analysis
Generalized Multicomponent spin model with arbitrary interactions network	Disciplinary productivity, defined in DEA in an input-output framework
node variable or multicomponent spin: $\mathbf{s}_i = (s_i^1, \dots, s_i^{(\gamma)}, \dots, s_i^N)$	deviation from world-country average productivity of discipline $i$ and countries ( $\gamma$ )
Pairwise node <i>interactions</i> or <i>couplings</i> : $J_{i,j}$	pairwise interdependencies between productivity of different disciplines
<i>Hamiltonian</i> : $H = -\frac{1}{2}\beta \sum_{i,j=1}^N J_{ij}\mathbf{s}_i(t) \cdot \mathbf{s}_j(t) - \sum_{i=1}^N \mathbf{s}_i(t) \cdot \mathbf{h}_i$	generalized cost function linked to the estimation of the likelihood
$\beta$ : inverse of the temperature	external global parameter
$\mathbf{h}_i$ : local magnetic field	contextual environmental variables of discipline $i$

The  $J_{ij}$  in physics measure a direct and reciprocal (mutual) effect (the interaction) of one entity on another entity (and vice versa). The concept of interaction in physics can find its correspondence in the interdependency in Input-Output economic analysis. The latter means the existence of a mutual influence between sectors (disciplines).

The coupling parameters  $J_{ij}$  generate the configurations of the system that may be characterized by the correlations between the spin variables, the so-called overlap measures (see also [63]), defined as follows:

$$Q_{ij} = 1/T \sum_{t=1}^T \mathbf{s}_i(t) \cdot \mathbf{s}_j(t), \quad (12)$$

where  $t = 1, \dots, T$  is time. As it is well known, a correlation measures the association between two variables. It shows a tendency of one variable to change with some regularity when the other changes, but this tendency may be moderated (influenced) by other factors, and depends on the whole configuration, including indirect effects. Correlation does not mean a direct effect or relation. On the other hand, interactions or interdependencies ( $J_{ij}$ ) refer to strict relationships between variables which allow us to describe the impact of the variation of one variable on another. Assuming this model to make inference permits us to consider correlations beyond the interdependencies among the units of analysis. As we will see in the application (see Section 5), the productivity of two disciplines may be correlated because they tend to be associated in their variation, but they may not interact. Here we impose  $J_{ij} \geq 0$  without loss of generality. The  $J_{ij}$  represent the interaction strength between  $i$  and  $j$ . The higher the value of  $J_{ij}$  the stronger is the interaction between  $i$  and  $j$ .

#### 4.3. Maximum-Likelihood and Pseudo-Likelihood Estimates

Though the likelihood function definition has deep roots in information theory, Bayesian inference and statistical mechanics, as discussed in the Section above, the realization of an optimization algorithm able to draw the optimal  $\{J\}$  is hindered by the general intractability of computing  $Z(\{J\})$  and its gradient [18,26]. Hence, in place of maximizing the likelihood function, we may define and maximize a different objective function, the so-called pseudo-likelihood. It is possible to show [18,64] that the estimation of the parameters obtained by a pseudo-likelihood maximization is consistent with the maximization of the likelihood function, that is the two functions are maximized by the same set of parameters. This statement becomes exact in the case of infinite sampling [18]. We do not discuss in detail here how these results are achieved. We discuss only the guidelines, redirecting for details elsewhere (see e.g., [65] and the references cited in this sub-section). By its very establishment, the pseudo-likelihood function permits to solve the optimization problem avoiding the troubles related to the computation of  $Z(\{J\})$ . It has indeed the advantage to be maximized in polynomial time. The pseudo-likelihood function is based on the so-called local conditional likelihood functions,

$p(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\})$  at each node of the network,  $\mathbf{s}_i, i = 1, \dots, N$ . The symbol  $\{\mathbf{s}_{\setminus i}\}$  means a set of variables  $\mathbf{s}_j$  with  $i \neq j$ . The local conditional probability (single variable pseudo-likelihood) at the  $i$ -th node is

$$p(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\}) = \frac{1}{Z_i} e^{-H_i(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\})} \tag{13}$$

The local Hamiltonian  $H_i(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\}) = -\mathbf{s}_i \cdot [\frac{1}{2} \sum_{i \neq j}^{1,N} J_{ij} \mathbf{s}_j]$  and the local partition function is  $Z_i = \sum_{\{\mathbf{s}_i\}} e^{-H_i(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\})}$ . Letting  $l'(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\}) = \log[p(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\})]$ , the Log-Pseudo-Likelihood function is

$$\lambda(\{J\}) = \sum_{t=1}^T \sum_{i=1}^N l'(\mathbf{s}_i(t)|\{\mathbf{s}_{\setminus i}(t)\}). \tag{14}$$

The gradient of the log-pseudo-likelihood function with respect to the parameter  $J_{ij}$  can be easily calculated, obtaining

$$\frac{\partial}{\partial J_{ij}} \lambda(\{J\}) = \frac{1}{2} T \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{s}_i(t) \cdot \mathbf{s}_j(t) - \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}} \right], \tag{15}$$

where  $\langle \cdot \rangle_{i,\{J\}}$  states for ensemble average calculated for the pdf  $p(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\})$  with parameters set  $\{J\}$ . It is possible to rephrase it, obtaining

$$\frac{\partial}{\partial J_{ij}} \lambda(\{J\}) = \frac{1}{2} T \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{s}_i(t) \cdot \mathbf{s}_j(t) - \langle \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}} \rangle_{\{J\}} \right], \tag{16}$$

where  $\langle \cdot \rangle_{\{J\}}$  states for ensemble average calculated with the pdf in Equation (7) or Equation (8) with set of parameters  $\{J\}$ . The gradient of the log likelihood function is

$$\frac{\partial}{\partial J_{ij}} l(\{J\}) = \frac{1}{2} T \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{s}_i(t) \cdot \mathbf{s}_j(t) - \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{\{J\}} \right] \tag{17}$$

By comparing Equations (15) and (16) it is possible to infer that in the limit of large  $T$ : (i) both the gradients go to zero if the elements of the set  $\{J\}$  are the ‘true’ parameters defining the pdf which generates the data; (ii)  $\frac{\partial}{\partial J_{ij}} \lambda(\{J\}) \rightarrow \frac{\partial}{\partial J_{ij}} l(\{J\})$ .

Specifics on computation of log-pseudo-likelihood function and its gradient with respect to  $J_{ij}$ , and details on optimization algorithm are reported in Appendix B.

The methods employed and the codes written to implement the related algorithms have been tested on the Ising model with known coupling coefficients on random graphs, thus guaranteeing the right convergence of the inference procedure and the proper reconstruction of the interaction network.

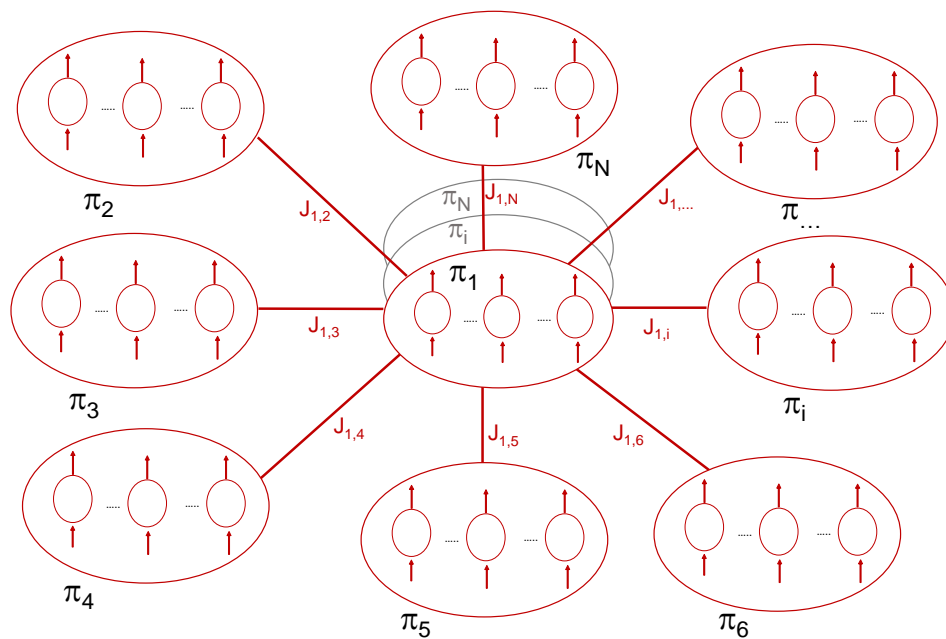
## 5. Application to Knowledge Production

### 5.1. The Knowledge Production Network

We point out that our proposed inferential approach used to recover the broader network structure can be applied to different network models developed for diverse fields of applications. To illustrate its potential, we apply the method to the field of knowledge production. Taking into account the data available for the empirical analysis, we will work with a network that has the form shown in Figure 3.

In the context of the GRFF model we assume that accumulated knowledge from previous periods is a fund variable and new knowledge produced in the form of publications in the current period is a flow. Two fund sources arise from accumulated knowledge: knowledge that a DMU itself produced in previous periods,  $z$ , and spillover knowledge accruing from the knowledge (publications) that other DMUs produced in previous periods,  $Y$ . In turn, the flow of new knowledge produced by a DMU in the current period becomes part of the fund of own accumulated knowledge it can draw upon in a

subsequent period and that new knowledge spills over and becomes a fund available to other DMUs in subsequent periods.



**Figure 3.** An illustration of the network model. Each disciplinary productivity  $\pi_i$ , which is a node in this network, includes the country-level set of disciplinary productivity  $\pi_i^\gamma$  with  $\gamma = 1, \dots, D$ .

Figure 3 illustrates the structure of the network that we will reconstruct using our new semi parametric approach in Section 5. This network models the interdependencies existing between disciplinary productivity/efficiency,  $\pi_i$ , where disciplines  $i = 1, \dots, N$  are the Scopus subject categories. The network nodes  $\pi_i$  comprise the respective productivity measures,  $\pi_i^1, \dots, \pi_i^\gamma, \dots, \pi_i^D$ , for each country  $\gamma = 1, \dots, D$ . We consider the main 53 world countries according to their scientific production. The productivity/efficiency of country  $\gamma$  in discipline  $i$ ,  $\pi_i^\gamma$  (omitting the time  $t$  from the notation for an easier reading), is computed in a DEA setup through the Shephard output distance function for each country relative to the discipline-time specific technology ( $P_i^{\gamma,t}$  that will be introduced in detail in Section 5) as

$$\pi_i^\gamma(x_i^{\gamma,t}, z_i^{\gamma,t}, Y_i^{\gamma,t}, y_i^{\gamma,t}) = \inf\{\pi_i^\gamma : y_i^{\gamma,t} / \pi_i^\gamma \in P_i^{\gamma,t}\}. \tag{18}$$

The reciprocal of the Shephard output distance function,  $1 / \pi_i^\gamma$ , measures the proportional expansion of observed outputs that could be achieved if the DMU were to become efficient.

Figure 3 shows a network of disciplinary productivities  $\pi_i$ , each of them composed by country-level productivity  $\pi_i^1, \dots, \pi_i^\gamma, \dots, \pi_i^D$ . For instance, the productivity of Chemistry  $\pi_{Chem}$  is composed by the country-level productivity  $\pi_{Chem}^{Arg}, \dots, \pi_{Chem}^\gamma, \dots, \pi_{Chem}^{USA}$  where ‘Arg’ stands for Argentina.

The disciplinary interdependencies or interactions are represented by the pathways,  $J_{ij}$ . Importantly, these pathways are generally unknown, and must be inferred, although if knowledge of the interactions is known, constraints can be incorporated that account for those interactions. Example disciplinary interdependencies include the use of new computational methods from computer science by those in mathematics or the natural sciences; advances in neuroscience by those in medicine and nursing; new findings in environmental science by those in earth sciences and agriculture. Indeed, as we will see in the application, the present study illustrates an interdependency pathway between physics and the social sciences. Although it is theoretically possible to include the inputs/outputs of all disciplines in the technology, by doing this we would encounter two problems. First, assuming the homogeneity of all disciplines in their knowledge production, which is clearly not the case, and second, the curse of dimensionality would likely render all DMUs to be on the efficient frontier.

Therefore, we first estimate productivity/efficiency for specific disciplines within a specific country in an NDEA model. Then, we use those productivity/efficiency estimates to infer the generally unknown connections between disciplines using the method described in Section 4.2.

### 5.2. Data and Descriptive Analysis

Our world-country bibliometric data were extracted from the Scopus database, for 16 disciplinary subject categories from 1996 to 2012. Data problems in bibliometric studies are well known. A common way to reduce them is to analyze macro-level bibliometric data. Comparative analysis is more reliable when the unit of analysis is more aggregated because in a larger sample size, micro random errors mutually compensate [66]. The potential for changes in coverage from inclusion or exclusion of journals to disproportionately affect smaller countries with fewer publications presents another issue of concern. This may lead to unreliable values when a country only has a small number of scholarly outputs [67]. To avoid this problem, we consider the 53 most productive countries (in terms of scientific productivity), which account for more than 95% of the world scientific production in the considered period. (The 53 countries (Country Alpha 3 Code labels) analysed in this paper are: 1 = ARG, 2 = AUS, 3 = AUT, 4 = BEL, 5 = BGR, 6 = BRA, 7 = CAN, 8 = CHE, 9 = CHL, 10 = CHN, 11 = COL, 12 = CZE, 13 = DEU, 14 = DNK, 15 = EGY, 16 = ESP, 17 = FIN, 18 = FRA, 19 = GBR, 20 = GRC, 21 = HKG, 22 = HRV, 23 = HUN, 24 = IND, 25 = IRL, 26 = IRN, 27 = ISR, 28 = ITA, 29 = JPN, 30 = KOR, 31 = MEX, 32 = MYS, 33 = NGA, 34 = NLD, 35 = NOR, 36 = NZL, 37 = PAK, 38 = POL, 39 = PRT, 40 = ROU, 41 = RUS, 42 = SAU, 43 = SGP, 44 = SRB, 45 = SVK, 46 = SVN, 47 = SWE, 48 = THA, 49 = TUN, 50 = TUR, 51 = TWN, 52 = UKR, 53 = USA).

Luwel [68] and Aksnes et al. [69] report the main issues related to the integration of bibliometric data with other inputs data, in particular R&D expenditures. Methodological problems in measuring productivity at the macro level are mainly due to a lack of standardization in the measurement of resources and outcomes across countries. Moreover, the methodologies for collecting input and output data have been developed largely independently from each other. We attempt to bypass these issues of standardization and measurement by working with simpler quantity data on inputs and outputs for this paper. We restrict research inputs to the number of publishing authors,  $NA$ , and outputs to the number of publications,  $P$ , as well as the number of highly cited publications (top 10%),  $HCP$ . These indicators are the most known and commonly used indicators for the assessment of the contributions scholars make in their research publications to the advancement of scholarly knowledge [70]. Bibliometrics and quantitative studies of science are heavily related to these indicators [71]. The data on  $NA$  come from Elsevier Bibliometric Research Project (EBRP). In the elaborations carried out to estimate the interactions ( $J_{ij}$ ) and infer the network structure, to increase the number of available data we transformed yearly data into weekly data by means of a linear interpolation. This leaves 833 observations of weekly total publications and highly cited publications for the 1996–2012 panel.

Table 2 gives an overview of all the indicators available for the present study. The list of the Scopus 27 subject categories is reported in Table 3.

**Table 2.** List of Indicators.

<b>Ind</b>	<b>Description</b>
<i>P</i>	Number of articles (integer count)
<i>P<sub>f</sub></i>	Number of articles (fractional count, based on authors affiliations)
<i>C</i>	Total citations (4 years window, i.e., for articles in 2006, citations are from 2006–2009)
<i>CPP</i>	Total citations per paper (4 years window, i.e., for articles in 2006, citations from 2006–2009)
<i>HCP</i>	Number of articles in top 10 per cent of most highly cited articles in a discipline
<i>PINT</i>	Number of internationally co-authored papers
<i>PNAT</i>	Number of nationally (but not internationally) co-authored papers
<i>PINST</i>	Number of papers co-authored by members of different institutions within a country
<i>PSA</i>	Number of non-collaborative (single address) papers
<i>NA</i>	Number of publishing authors

**Table 3.** List of the 27 Scopus' subject categories.

<b>asjc Code</b>	<b>Subject Category</b>	<b>Description</b>
10	GENE	General
11	<b>AGRI</b>	Agricultural and Biological Sciences
12	ARTS	Arts and Humanities
13	<b>BIOC</b>	Biochemistry, Genetics and Molecular Biology
14	BUSI	Business, Management and Accounting
15	<b>CENG</b>	Chemical Engineering
16	<b>CHEM</b>	Chemistry
17	<b>COMP</b>	Computer Science
18	DECI	Decision Sciences
19	<b>EART</b>	Earth and Planetary Sciences
20	ECON	Economics, Econometrics and Finance
21	<b>ENER</b>	Energy
22	<b>ENGI</b>	Engineering
23	<b>ENVI</b>	Environmental Science
24	<b>IMMU</b>	Immunology and Microbiology
25	<b>MATE</b>	Materials Science
26	<b>MATH</b>	Mathematics
27	<b>MEDI</b>	Medicine
28	<b>NEUR</b>	Neuroscience
29	NURS	Nursing
30	<b>PHAR</b>	Pharmacology, Toxicology and Pharmaceutics
31	<b>PHYS</b>	Physics and Astronomy
32	PSYC	Psychology
33	SOCI	Social Sciences
34	VETE	Veterinary
35	DENT	Dentistry
36	HEAL	Health Professions

Table 3 reports in bold the 16 subject categories considered in the analysis and the variables available for each discipline. We excluded social sciences and humanities disciplines whose coverage of scholarly outputs in indexed journals is much lower than the other subject categories considered. Table 4 presents descriptive statistics for the selected disciplines, namely: BIOC, COMP, ENGI, MEDI and PHYS. Appendix C reports descriptive statistics on the other disciplines that will be analysed in the paper.

**Table 4.** Summary Statistics Performance Variables, by Discipline (ASJC code) (53 Countries, 1996–2012, 901 Observations Each Discipline).

<b>BIOC (13)</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
Articles	4357.6	9723.4	44	85,295
Highly Cited	529.1	1572.1	0	15,480
Number of Authors	10,499.4	23,931.7	95	229,139
<b>COMP (17)</b>				
Articles	2727.5	7315.9	7	80,276
Highly Cited	355.6	1013.9	0	10,123
Number of Authors	4161.6	11,548.7	11	128,273
<b>ENGI (22)</b>				
Articles	5433.7	14,078.9	31	156,349
Highly Cited	739.5	1804.5	0	19,830
Number of Authors	8981.8	23,456.8	54	293,605
<b>MEDI (27)</b>				
Articles	7672.4	17,175.3	95	165,181
Highly Cited	1027.9	2929.3	0	28,743
Number of Authors	15,773.2	34,286.3	211	351,702
<b>PHYS (31)</b>				
Articles	4396.5	8271.2	19	58,244
Highly Cited	559.9	1208.5	0	10,591
Number of Authors	6946.0	14,391.0	26	127,209

### 5.3. Models for Estimating Knowledge Production

Our network models borrow from previous work on knowledge production ([6,30]), while the production variable choices are modeled after Georgescu-Roegen, where we include both flow variables and funds variables. We think of knowledge production in a production axiomatic framework, in which we include author count as a flow type input variable, and cumulated previous own publications as a fund or knowledge stock, which produces a flow of publication outputs.

What makes it a network is the fact that the cumulated publications of a discipline in one country are available to other countries in the same discipline, and previous cumulated publications from other countries in that discipline also are available to the discipline in the ‘home’ country as a fund-type input variable. This specification proxies the public good/externality nature of publications as well as their role in contributing to the stock of knowledge.

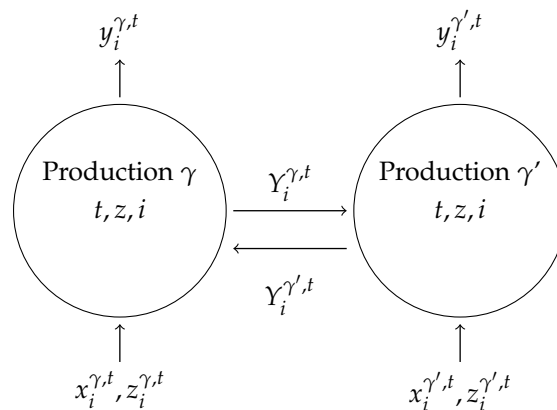
Table 5 outlines the examined basic (static) and network productivity models, which we separate into two categories: Quantity (1) and Quality (2). The first category represents ‘quantity of knowledge,’ using raw publication counts ( $P$ ) to measure final knowledge output. Within this category, the basic model (1.1) uses the stock of previous publications within each country and the number of authors ( $NA$ ) as the knowledge inputs. The corresponding network model (1.2) also includes the stock of previous publications from other countries as a knowledge spillover input. The second category represents ‘quality of knowledge,’ using the number of highly cited publications ( $HCP$ ) in place of raw publication counts, for both previous publication inputs and current publication final outputs. While citations are widely used to indicate publication quality, we note that there are several other potential alternatives, including SCImago journal rank (SJR) (Lee et al., 2016), peer reviews [72], and restricting to top-tier journals [73]. Quality models 2.1 and 2.2 serve as the quality-adjusted versions of quantity models 1.1 and 1.2, using the HCP approach to distinguish quality. We add a simple productivity model given by the ratio of publications to authors ( $P/NA$ ) as a baseline.



**Table 5.** Network DEA Models of Knowledge Production.

Models	Inputs	Output
<b>(0) Simple productivity model</b>	Own author count ( $NA$ )	Own current pubs ( $P$ )
<b>(1) Quantity</b>		
1.1 (Basic)	$NA$ , own prev pubs	$P$
1.2 (Network)	$NA$ , own prev pubs, other prev pubs	$P$
<b>(2) Quality</b>		
2.1 (Basic)	$NA$ , own prev $HCP$	Own current $HCP$
2.2 (Network)	$NA$ , own prev $HCP$ , other prev $HCP$	Own current $HCP$

Figure 4 illustrates the network for the hypothetical case of two countries,  $\gamma$  and  $\gamma'$ , both for discipline  $i$  in period  $t$ . Their flow inputs are the author counts denoted by  $x$  and their final outputs are denoted by  $y$ . The previous publication fund variables that provide the network connection are denoted as  $Y$  and the own previous fund variables are denoted by  $z$ .



**Figure 4.** Network Technology for Knowledge Production.

5.4. Production Efficiency Estimated Using NDEA

To fix ideas, and following [6], let each country be indexed by  $\gamma = 1, \dots, D$ , for periods  $t = 1, \dots, T$ . We will augment their model by including disciplines  $i = 1, \dots, N$ , here  $N = 16$ . Denote flow input as  $x^{\gamma,t}$  (here a scalar, but possibly a vector). Fund variables include own country cumulated past publications denoted as  $z_i^{\gamma,t}$ , and other country cumulated previous publications as  $Y_i^{\gamma,t}$ . Final output is own country current period publications denoted as  $y_i^{\gamma,t}$ .

Again, following [6], we define the own country fund variable  $z$  as the sum of the previous 3 periods' publications for that country, which we denote as  $z_i^{\gamma,t} = \sum_{\tau=1}^3 y_i^{\gamma,t-\tau}$ , where  $\tau = 1, 2, 3$  represents the 3 previous years. Similarly, we use  $Y_i^{\gamma,t} = \sum_{\tau=1}^3 \sum_{\gamma' \neq \gamma}^D y_i^{\gamma',t-\tau}$  to represent spillover knowledge from other countries' previous publications.

We can now state the formal problem we are estimating to solve for efficiency in the NDEA model. As explained in Section 5.1, we employ the Shephard output distance function as a performance measure for each country, i.e., it is the explicit objective function. The distance function for country  $\gamma$  in discipline  $i$ ,  $\pi_i^\gamma$ , is defined as

$$\pi_i^\gamma(x_i^{\gamma,t}, z_i^{\gamma,t}, Y_i^{\gamma,t}, y_i^{\gamma,t}) = \inf\{\pi_i^\gamma : y_i^{\gamma,t} / \pi_i^\gamma \in P_i^{\gamma,t}(x_i^{\gamma,t}, z_i^{\gamma,t}, Y_i^{\gamma,t})\}. \tag{19}$$

This objective function scales observed country output to the frontier of the output set  $P^{\gamma,t}(\cdot)$  and takes a value of unity for a country on the frontier and a value less than one for a country that is below the frontier. Using DEA (see Section 3.1), the reference technology (output set) for period  $t$ , serves as the constraints in our problem:

$P^t(x_i^{\gamma,t}, z_i^{\gamma,t}, Y_i^{\gamma,t}) = \{y : (x_i^{\gamma,t}, z_i^{\gamma,t}, Y_i^{\gamma,t}) \text{ can produce } y\}$  may be written

$$P_i^{\gamma,t}(x_i^{\gamma,t}, z_i^{\gamma,t}, Y_i^{\gamma,t}) = \{y^t : \sum_{\gamma=1}^D \lambda_i^{\gamma,t} y_i^{\gamma,t} \geq y, \quad (20)$$

$$\sum_{\gamma=1}^D \lambda_i^{\gamma,t} x_i^{\gamma,t} \leq x_i^{\gamma,t}, \quad (21)$$

$$\sum_{\gamma=1}^D \lambda_i^{\gamma,t} z_i^{\gamma,t} \leq z_i^{\gamma,t} \quad (22)$$

$$\sum_{\gamma=1}^D \lambda_i^{\gamma,t} Y_i^{\gamma,t} \leq Y_i^{\gamma,t}, \quad (23)$$

$$\lambda_i^{\gamma,t} \geq 0, \gamma = 1, \dots, D; i = 1, \dots, N; t = 1, \dots, T, \}$$

where the  $\lambda_i^{\gamma,t}$  are intensity variables that form the best-practice frontier technology from the observed inputs and outputs. Equation (20) is the constraint with respect to own country current output (publications), Equation (21) is the input flow constraint, Equation (22) is a fund constraint with respect to own cumulated previous publications and Equation (23) is a fund constraint with respect to other country cumulated publications—the spillover constraint. We solve this problem for each discipline in each country using linear programming. These solutions are the data for the second stage which estimates the correlations Q and the parameters J, as described below.

As described in Section 4.2, the statistical analysis of the knowledge production we propose here is developed in the framework of information theory, where through the well-known Shannon theorem, the entropy is introduced and defined. We can measure the so-called pairwise correlation functions or overlaps from the collected data. Assuming that we only get this information, by following the line of the Shannon theorem it is possible to find the class of the probability distribution (which will depend on some adjustable parameters, the set J), which maximizes the entropy with the constraint that the pairwise correlations obtained from such probability distribution match the measured ones. In this way the probability distribution (see Equation (12) with the Hamiltonian of the so-called Ising model (see Equation (13)), and also Appendix A), is directly obtained by maximizing the entropy introduced in the Shannon theorem with the aforementioned constraint, without any further unverifiable (thus arbitrary) hypothesis. In this respect, the entropy introduced in the Shannon theorem represents a measure of the ‘amount of uncertainty’. For example, limiting to the case of a discrete probability distribution, if any constraint is superimposed (i.e., any a priori knowledge is available), the entropy function introduced by Shannon is maximum for a uniform distribution. This agrees with our intuitive notion that a broad distribution represents more uncertainty than a sharply peaked one. Figure 3 is a sketch of an Ising model whose variables are defined in the knowledge production framework. It is thus linked to the entropy defined from information theory because the adoption of the Ising model is a direct consequence of the maximum entropy principle, as discussed above.

### 5.5. Results

We estimate our two models for each of the 53 countries and 16 disciplines, reported in Table 3. To summarize the results, we present the annual output share-weighted geometric means of the productivity values by discipline, for the alternate versions of Models 1 and 2: Total publications (basic and network, 1.1 and 1.2) and highly cited publications (basic and network, 2.1 and 2.2). For comparison with our model estimates we include the simple ratio of publications to authors ( $P/NA$ ). The rank order correlations of  $P/NA$  and estimated production efficiency are reported in Table 6 for selected disciplines: Computer Science (COMP), Engineering (ENGI), Medicine (MEDI) and Physics (PHYS). Not surprisingly, the lowest correlations occur for the basic DEA quantity model (1.1) and the NDEA quality model (2.2). In contrast, the two quantity models (1.1 and 1.2) and the two quality models (2.1 and 2.2) tend to have the highest correlations in each of the disciplines.

**Table 6.** Rank Correlations of Productivity for Selected Disciplines (COMP, ENGI, MEDI and PHYS).

DEA and NDEA Models						
COMP	<i>P/NA</i>	1.1	1.2	2.1	2.2	
<i>P/NA</i>	1.000					
1.1	0.532	1.000				
1.2	0.389	0.635	1.000			
2.1	0.417	0.314	0.262	1.000		
2.2	0.229	0.104	0.416	0.719	1.000	
ENGI	<i>P/NA</i>	1.1	1.2	2.1	2.2	
<i>P/NA</i>	1.000					
1.1	0.545	1.000				
1.2	0.437	0.646	1.000			
2.1	0.386	0.430	0.337	1.000		
2.2	0.139	0.222	0.555	0.655	1.000	
MEDI	<i>P/NA</i>	1.1	1.2	2.1	2.2	
<i>P/NA</i>	1.000					
1.1	0.605	1.000				
1.2	0.548	0.805	1.000			
2.1	0.544	0.413	0.418	1.000		
2.2	0.350	0.279	0.502	0.729	1.000	
PHYS	<i>P/NA</i>	1.1	1.2	2.1	2.2	
<i>P/NA</i>	1.000					
1.1	0.536	1.000				
1.2	0.165	0.371	1.000			
2.1	0.390	0.382	0.235	1.000		
2.2	−0.032	−0.033	0.452	0.619	1.000	

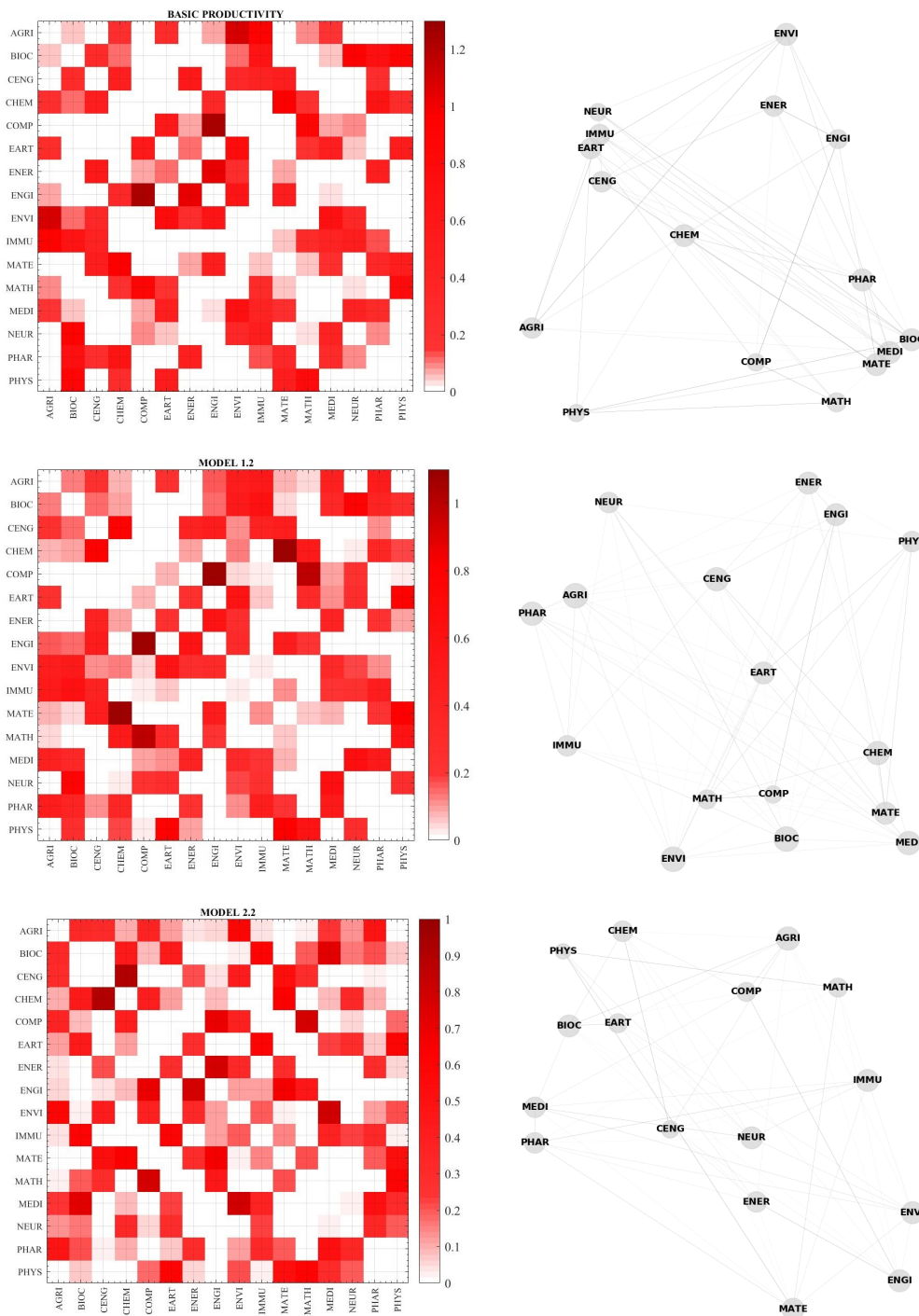
We are interested in studying the interdependencies across disciplines as well as within disciplines at a macro or country level. The interactions between knowledge production efficiency may differ by type and magnitude, as well as their ultimate effects on the associated disciplines and countries. Our NDEA estimates provide within disciplinary connection across countries, but not across disciplines.

For instance, collaboration between researchers in two disciplines, say physics and health can mean that gains in efficiency in one discipline are reinforced by gains in efficiency in another discipline. Interdisciplinary research requires that the researchers in each discipline to learn the vocabulary, terminology, the notation, the ideas, and so forth of other disciplines. Those disciplines and researchers who can overcome the transaction costs of learning the vocabulary, terminology, etc. can expand this more general/interdisciplinary knowledge. Our method identifies those disciplines where such gains are possible. In contrast, if interdisciplinary transaction costs are too high, as might occur when researchers come from two very different disciplines, then gains in efficiency in one discipline may serve to lower efficiency in another discipline.

There may also be cases in which two or more disciplines with different levels of production efficiency interact such that the more productive disciplines slow down while the less productive disciplines increase their productivity. For this reason, analyzing the correlations or indirect connections (the overlap measures  $Q_{ij}$  introduced in Equation (12)) between (disciplinary) efficiency levels can be interesting. However, going further and estimating their interdependencies ( $J_{ij}$ ) can provide more useful information to analyze the way in which scientific knowledge is produced and organized worldwide. This information could be useful for policy makers who determine which disciplines or topics to prioritize and how to distribute research funds among disciplines.

Figure 5 shows the estimated interdependency parameters ( $J_{ij}$ ) -left panels- and the inferred networks -right panels for the three production efficiency models. Top panels refer to the simple productivity model (*P/NA*), middle panels show the Model 1.2 (network quantity model) and the

bottom panels report Model 2.2 (network quality model) results. In the left panels of Figure 5, the darker squares indicate higher  $J_{ij}$  and the NW to SE diagonal comprises all white squares since  $J_{ii} = 0$  by hypothesis.



**Figure 5.** Estimated  $J_{ij}$  (left panels) and inferred networks (right panels) for the three production efficiency models. Top panels refer to the simple productivity model ( $P/NA$ ), middle panels show the network quantity model 1.2 and the bottom panels show the network quality model 2.2 results.

The reconstructed networks reported in the right panels of Figure 5 are derived from the estimated  $J_{ij}$  obtained by the maximization of the pseudo-likelihood function. The  $J_{ij}$  are the edges. The diameter

of the node for discipline  $i$  is proportional to the number of interactions  $J_{ij}$ . The thickness of the edge depends on the intensity of the related interaction.

Figure 6 shows the calculated overlap measures ( $Q_{ij}$ ) and the estimated interdependencies ( $J_{ij}$ ) for the different productivity models. The methodology introduced in Section 4 hence allowed us to empirically infer the network structure existing among disciplinary productivity going beyond the simple overlap (or correlations) measures ( $Q_{ij}$ ).

	AGRI	BIOC	CENG	CHEM	COMP	EART	ENER	ENGI	ENVI	IMMU	MATE	MATH	MEDI	NEUR	PHAR	PHYS
AGRI	-	<b>0.055</b>	<b>0.000</b>	<b>0.210</b>	<b>0.000</b>	<b>0.256</b>	<b>0.000</b>	<b>0.067</b>	<b>1.080</b>	<b>0.895</b>	<b>0.000</b>	<b>0.084</b>	<b>0.162</b>	<b>0.012</b>	<b>0.000</b>	<b>0.004</b>
BIOC	0.182	-	<b>0.293</b>	<b>0.119</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.102</b>	<b>0.650</b>	<b>0.000</b>	<b>0.000</b>	<b>0.060</b>	<b>0.907</b>	<b>0.695</b>	<b>0.874</b>	
CENG	0.147	0.167	-	<b>0.466</b>	<b>0.000</b>	<b>0.000</b>	<b>0.584</b>	<b>0.000</b>	<b>0.318</b>	<b>0.422</b>	<b>0.472</b>	<b>0.004</b>	<b>0.000</b>	<b>0.003</b>	<b>0.233</b>	<b>0.000</b>
CHEM	0.141	0.144	0.132	-	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.316</b>	<b>0.000</b>	<b>0.000</b>	<b>0.970</b>	<b>0.188</b>	<b>0.000</b>	<b>0.000</b>	<b>0.645</b>	<b>0.253</b>
COMP	0.148	0.145	0.121	0.111	-	<b>0.551</b>	<b>0.068</b>	<b>1.223</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.867</b>	<b>0.077</b>	<b>0.089</b>	<b>0.000</b>	<b>0.000</b>
EART	0.167	0.149	0.122	0.106	0.163	-	<b>0.118</b>	<b>0.000</b>	<b>0.768</b>	<b>0.000</b>	<b>0.000</b>	<b>0.166</b>	<b>0.467</b>	<b>0.053</b>	<b>0.000</b>	<b>0.518</b>
ENER	0.140	0.144	0.136	0.116	0.139	0.134	-	<b>1.025</b>	<b>0.236</b>	<b>0.000</b>	<b>0.072</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.460</b>	<b>0.000</b>
ENGI	0.130	0.136	0.120	0.106	0.155	0.086	0.143	-	<b>0.613</b>	<b>0.000</b>	<b>0.481</b>	<b>0.000</b>	<b>0.038</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
ENVI	0.182	0.176	0.149	0.125	0.145	0.177	0.140	0.127	-	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.690</b>	<b>0.333</b>	<b>0.002</b>	<b>0.000</b>
IMMU	0.193	0.206	0.161	0.156	0.144	0.162	0.142	0.130	0.181	-	<b>0.042</b>	<b>0.302</b>	<b>0.401</b>	<b>0.481</b>	<b>0.135</b>	<b>0.000</b>
MATE	0.179	0.174	0.156	0.158	0.152	0.145	0.143	0.141	0.167	0.191	-	<b>0.060</b>	<b>0.212</b>	<b>0.000</b>	<b>0.339</b>	<b>0.491</b>
MATH	0.098	0.100	0.102	0.089	0.134	0.144	0.094	0.080	0.097	0.111	0.119	-	<b>0.000</b>	<b>0.034</b>	<b>0.013</b>	<b>0.775</b>
MEDI	0.218	0.220	0.157	0.169	0.174	0.193	0.153	0.154	0.210	0.254	0.213	0.103	-	<b>0.425</b>	<b>0.269</b>	<b>0.000</b>
NEUR	0.158	0.163	0.145	0.130	0.137	0.145	0.119	0.099	0.156	0.179	0.150	0.098	0.203	-	<b>0.084</b>	<b>0.000</b>
PHAR	0.152	0.171	0.144	0.136	0.115	0.105	0.125	0.119	0.147	0.166	0.151	0.084	0.173	0.132	-	<b>0.000</b>
PHYS	0.111	0.093	0.092	0.067	0.114	0.171	0.095	0.071	0.116	0.100	0.122	0.127	0.080	0.087	0.064	-

	AGRI	BIOC	CENG	CHEM	COMP	EART	ENER	ENGI	ENVI	IMMU	MATE	MATH	MEDI	NEUR	PHAR	PHYS
AGRI	-	<b>0.127</b>	<b>0.234</b>	<b>0.083</b>	<b>0.000</b>	<b>0.238</b>	<b>0.000</b>	<b>0.158</b>	<b>0.458</b>	<b>0.508</b>	<b>0.080</b>	<b>0.051</b>	<b>0.414</b>	<b>0.000</b>	<b>0.471</b>	<b>0.000</b>
BIOC	0.151	-	<b>0.150</b>	<b>0.098</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.144</b>	<b>0.485</b>	<b>0.602</b>	<b>0.049</b>	<b>0.000</b>	<b>0.344</b>	<b>0.762</b>	<b>0.362</b>	<b>0.249</b>
CENG	0.119	0.169	-	<b>0.785</b>	<b>0.000</b>	<b>0.000</b>	<b>0.410</b>	<b>0.480</b>	<b>0.112</b>	<b>0.410</b>	<b>0.453</b>	<b>0.000</b>	<b>0.000</b>	<b>0.003</b>	<b>0.104</b>	<b>0.000</b>
CHEM	0.141	0.192	0.182	-	<b>0.000</b>	<b>0.000</b>	<b>0.098</b>	<b>0.000</b>	<b>0.125</b>	<b>0.000</b>	<b>1.081</b>	<b>0.495</b>	<b>0.000</b>	<b>0.027</b>	<b>0.357</b>	<b>0.186</b>
COMP	0.115	0.161	0.144	0.139	-	<b>0.081</b>	<b>0.002</b>	<b>1.066</b>	<b>0.039</b>	<b>0.034</b>	<b>0.000</b>	<b>0.995</b>	<b>0.097</b>	<b>0.210</b>	<b>0.000</b>	<b>0.033</b>
EART	0.088	0.115	0.080	0.098	0.120	-	<b>0.227</b>	<b>0.000</b>	<b>0.572</b>	<b>0.062</b>	<b>0.000</b>	<b>0.326</b>	<b>0.110</b>	<b>0.267</b>	<b>0.000</b>	<b>0.783</b>
ENER	0.085	0.145	0.153	0.143	0.134	0.096	-	<b>0.583</b>	<b>0.252</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.385</b>	<b>0.006</b>	<b>0.214</b>	<b>0.096</b>
ENGI	0.129	0.187	0.179	0.160	0.211	0.100	0.170	-	<b>0.306</b>	<b>0.000</b>	<b>0.475</b>	<b>0.205</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
ENVI	0.133	0.168	0.126	0.134	0.140	0.130	0.125	0.162	-	<b>0.020</b>	<b>0.000</b>	<b>0.000</b>	<b>0.326</b>	<b>0.174</b>	<b>0.104</b>	<b>0.000</b>
IMMU	0.135	0.207	0.152	0.167	0.133	0.103	0.129	0.137	0.138	-	<b>0.109</b>	<b>0.000</b>	<b>0.198</b>	<b>0.227</b>	<b>0.459</b>	<b>0.000</b>
MATE	0.145	0.205	0.188	0.223	0.168	0.098	0.155	0.199	0.145	0.178	-	<b>0.059</b>	<b>0.085</b>	<b>0.000</b>	<b>0.202</b>	<b>0.792</b>
MATH	0.081	0.136	0.109	0.153	0.174	0.108	0.119	0.170	0.108	0.094	0.167	-	<b>0.000</b>	<b>0.000</b>	<b>0.583</b>	
MEDI	0.166	0.234	0.181	0.189	0.171	0.117	0.165	0.191	0.186	0.217	0.216	0.134	-	<b>0.627</b>	<b>0.482</b>	<b>0.000</b>
NEUR	0.135	0.177	0.146	0.160	0.131	0.093	0.105	0.140	0.129	0.160	0.158	0.097	0.181	-	<b>0.000</b>	<b>0.245</b>
PHAR	0.135	0.177	0.134	0.163	0.105	0.070	0.122	0.155	0.128	0.139	0.159	0.100	0.181	0.144	-	<b>0.000</b>
PHYS	0.087	0.130	0.096	0.127	0.137	0.132	0.110	0.126	0.095	0.103	0.151	0.144	0.115	0.106	0.081	-

	AGRI	BIOC	CENG	CHEM	COMP	EART	ENER	ENGI	ENVI	IMMU	MATE	MATH	MEDI	NEUR	PHAR	PHYS
AGRI	-	<b>0.323</b>	<b>0.299</b>	<b>0.105</b>	<b>0.346</b>	<b>0.115</b>	<b>0.039</b>	<b>0.051</b>	<b>0.586</b>	<b>0.047</b>	<b>0.000</b>	<b>0.022</b>	<b>0.248</b>	<b>0.135</b>	<b>0.475</b>	<b>0.000</b>
BIOC	0.127	-	<b>0.000</b>	<b>0.425</b>	<b>0.082</b>	<b>0.432</b>	<b>0.011</b>	<b>0.008</b>	<b>0.026</b>	<b>0.594</b>	<b>0.000</b>	<b>0.197</b>	<b>0.729</b>	<b>0.164</b>	<b>0.213</b>	<b>0.075</b>
CENG	0.097	0.081	-	<b>0.896</b>	<b>0.000</b>	<b>0.000</b>	<b>0.210</b>	<b>0.036</b>	<b>0.407</b>	<b>0.014</b>	<b>0.516</b>	<b>0.290</b>	<b>0.000</b>	<b>0.000</b>	<b>0.018</b>	<b>0.000</b>
CHEM	0.109	0.125	0.134	-	<b>0.395</b>	<b>0.116</b>	<b>0.000</b>	<b>0.094</b>	<b>0.000</b>	<b>0.000</b>	<b>0.640</b>	<b>0.000</b>	<b>0.080</b>	<b>0.321</b>	<b>0.108</b>	<b>0.000</b>
COMP	0.140	0.135	0.104	0.137	-	<b>0.000</b>	<b>0.000</b>	<b>0.691</b>	<b>0.387</b>	<b>0.000</b>	<b>0.000</b>	<b>0.766</b>	<b>0.000</b>	<b>0.047</b>	<b>0.010</b>	<b>0.172</b>
EART	0.118	0.120	0.078	0.075	0.135	-	<b>0.275</b>	<b>0.000</b>	<b>0.000</b>	<b>0.603</b>	<b>0.000</b>	<b>0.000</b>	<b>0.223</b>	<b>0.286</b>	<b>0.077</b>	<b>0.600</b>
ENER	0.081	0.097	0.077	0.083	0.092	0.067	-	<b>0.772</b>	<b>0.325</b>	<b>0.000</b>	<b>0.289</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.287</b>	<b>0.050</b>
ENGI	0.092	0.104	0.093	0.127	0.155	0.071	0.115	-	<b>0.114</b>	<b>0.114</b>	<b>0.665</b>	<b>0.462</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
ENVI	0.122	0.138	0.088	0.107	0.141	0.108	0.084	0.114	-	<b>0.188</b>	<b>0.026</b>	<b>0.000</b>	<b>0.788</b>	<b>0.000</b>	<b>0.121</b>	<b>0.216</b>
IMMU	0.108	0.137	0.073	0.091	0.114	0.108	0.085	0.090	0.123	-	<b>0.147</b>	<b>0.000</b>	<b>0.367</b>	<b>0.227</b>	<b>0.322</b>	<b>0.029</b>
MATE	0.102	0.119	0.142	0.141	0.134	0.097	0.113	0.144	0.119	0.113	-	<b>0.216</b>	<b>0.000</b>	<b>0.000</b>	<b>0.194</b>	<b>0.524</b>
MATH	0.108	0.105	0.112	0.098	0.141	0.102	0.077	0.102	0.113	0.070	0.121	-	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.596</b>
MEDI	0.126	0.142	0.088	0.104	0.124	0.110	0.085	0.098	0.146	0.129	0.118	0.106	-	<b>0.025</b>	<b>0.503</b>	<b>0.303</b>
NEUR	0.094	0.107	0.082	0.117	0.114	0.063	0.072	0.073	0.103	0.099	0.076	0.063	0.102	-	<b>0.331</b>	<b>0.201</b>
PHAR	0.117	0.125	0.103	0.120	0.118	0.071	0.085	0.103	0.121	0.118	0.120	0.074	0.136	0.119	-	<b>0.000</b>
PHYS	0.105	0.101	0.072	0.073	0.119	0.110	0.072	0.081	0.106	0.078	0.099	0.112	0.091	0.061	0.054	-

Figure 6. Overlaps ( $Q_{ij}$ ) and Interdependencies ( $J_{ij}$ ) of the simple productivity model ( $P/NA$ ) (top panel), the network quantity model 1.2 (middle panel) and the network quality model 2.2 (bottom panel). The Northeast values reported in bold are the  $J_{ij}$  while the Southwest values correspond to the  $Q_{ij}$ . See Table 5 for model specifications.

For instance, we can analyze the correlations and interdependencies between the productivity of disciplines CHEM with IMMU and MATE. In the basic quantity model (Figure 6—top panel) CHEM and IMMU and CHEM and MATE show the same overlap measure  $Q_{ij} = 0.16$ , meaning that their production efficiency tend to be positively associated. On the other hand, their respective interdependencies are different. In fact,  $J_{ij}$  between CHEM and IMMU is zero, while the interdependency between CHEM and MATE is 0.970, meaning that the productivities of CHEM and MATE present a high level of interdependency (mutual interaction). Similarly, the correlations (indirect connection) between PHYS with IMMU and MATE are respectively 0.10 and 0.12, while their interdependencies are respectively 0 and 0.49. This means that PHYS interacts with MATE, but not with IMMU, although the respective production efficiencies are correlated.

Inspecting the middle and the bottom panels of Figure 6 we note that the interaction ( $J_{ij}$ ) between PHYS and MATE is 0.79 in the quantity model (1.2) and 0.52 in the quality model (2.2) while the respective correlations are 0.15 and 0.1.

We can conduct a comparative qualitative analysis between the three different models to explain the results and how the technique works. We solved an indirect problem to estimate the  $J_{ij}$  values for each dataset and each model separately, so the inferred network structure is the optimal structure for each dataset.

For instance, the values of  $J_{AGRI,PHAR}$  in the three panels of Figure 6 are 0.000 (top), 0.471 (middle) and 0.475 (bottom). These values indicate there is no interdependency between AGRI and PHAR in the simple productivity model ( $P/NA$ ) as the score is zero, while the interdependencies between AGRI and PHAR are much stronger in the quantity and quality models in which the knowledge production assumes as inputs  $NA$ , own previous publications and other previous publications and as output own current publications (Mod. 1.2) while Model 2.2 considers the same inputs/output than model 1.2 but uses the number of Highly Cited Publications ( $HCP$ ). Similarly,  $J_{MATH,PHAR}$  is 0.013 according to the simple productivity model, showing a weak interaction among the two simple productivities estimated by the number of publications per author, while there is no interaction between the productivities of the two disciplines if we measure them according to Model 1.2 and 2.2 (see Table 5) as the  $J$  values for MATH and PHAR in middle and bottom panels are zero.

We can also compare the simple productivity model ( $P/NA$ ) with the quantity models (1.1 and 1.2) to consider any possible effects posed by the curse of dimensionality, given our sample size and model dimensions. The overall results are quite similar: the only differences are due to the DEA modelling and so we should prefer the results of the Model 1.2 which accounts for the Georgescu-Roegen's fund's modelling of the knowledge production.

We then compared the quantity vs. quality results for the network models (1.2 and 2.2), to see whether the obtained estimates were consistent. We found for instance that the interaction between Physics (PHYS) and Computer Science (COMP) in the quantity model (1.2) is 0.033 while their interaction in the quality model (2.2) increases up to 0.172. In contrast, the interaction between Physics and Chemistry that is 0.186 in the quantity model goes down to zero in the quality model.

As we may expect, the interactions between CHEM and CENG are quite high in all the three models (0.466 in the simple productivity model, 0.785 in Mod. 1.2 and 0.896 in Mod 2.2); other expected results are the high interactions of COMP with ENGI and MATH because these disciplines share the same community. A striking result is the interactions we observe between PHYS and MEDI which is quite high in the Quality model (Mod. 2.2, with a value of 0.303) but absent in the Quantity model (Mod. 1.2). A policy implication of this result could be made in the discussion about supporting Societal Challenges that are focused on Medical Sciences and the important role that Physics could play in this context.

The results commented in this section show the usefulness of the analyses carried out in the previous section to shed deeper and new light on the interactions among disciplinary productivity that we would not have been able to derive if we had reduced the analysis to the simple productivity model ( $P/NA$ ). Differently from existing bibliometric literature [74,75] we consider not only the outputs of scientific production, but also their efficiency in knowledge production. In addition, differently

from existing efficiency literature [6] we estimate the interdependencies among disciplinary efficiency thanks to the application of the inferential approach proposed in Section 4.

## 6. Discussion and Conclusions

Network Data Envelopment Analysis (NDEA) models are often used to measure producer performance when separate production technologies are linked across divisions, time, and spillovers can occur between different producers. In theory, NDEA models can accommodate complex, multi-product technologies with inputs used to produce intermediate products and final outputs. However, in practice, the degree of complexity a researcher can introduce is limited by the number of available observations used to construct the technology in what is known as the curse of dimensionality. It is also not obvious how to test for the appropriate structure of the network. In this paper we address these issues with a new theoretical method based on Shannon's entropy that allows us to infer wider linkages between various producers without having to specify those links within the NDEA model. Our method specifies a NDEA technology and provides nonparametric estimates of producer performance relative to that assumed technology. Then, in a second stage, we employ a semiparametric Bayesian framework that allows us to estimate, rather than assume, the network structure. This second stage exploits advances in the physics of complex systems, machine learning and econometrics of information and reveals additional linkages in the network—in our case—allowing us to infer connections between knowledge disciplines.

While we consider our main contribution to be providing an inferential method to identify structure in NDEA models, we consider our application to knowledge production to be of interest in its own right. The economics of science [76] reminds us that researchers do research for different reasons, including their interest in “puzzle solving”, reputation based on the priority of their discovery, awards and recognition for their achievements, and also through publications which can play a key role in funding and promotion. Research is also a public good, generating knowledge spillovers that can be difficult to capture and quantify. Like any public good, this can lead to underprovision in the market.

The economics of science tells us that the production of scientific research involves multiple inputs, including knowledge, time, materials and equipment. Some inputs are embedded in people (knowledge and time in particular), and most of these inputs are expensive. As observed by Stephan [76], incentives and cost matter for science and economics, particularly for shaping the most efficient mix of resource allocation across disciplines.

In this paper, we combine concepts from economic production theory, Bayesian statistics, and the physics of complex systems to infer the cross-disciplinary and cross-country interactions of research activities. Such understanding is key to achieving the efficient mix of resources for research. The NDEA estimates can be used to derive correlations between disciplines and those same NDEA estimate can be used in a second stage to infer interdependencies between disciplines. For instance, controlling for the quality of publications, math and medicine exhibit positive association, but zero interdependency. In other cases, such as physics and math, there are relatively low correlations of productive efficiencies between the two disciplines, but a high interdependency. We find non-trivial interactions in many cases, which seems promising for future work in this area. Our framework and results could be of particular interest to policy-makers and agencies tasked with prioritizing research funding areas. For instance, the relatively high interaction between physics and medicine might suggest the need to include topics from the physical sciences in new funding for medical research.

This approach could also be applied to other knowledge network systems beyond academic research, such as innovation and technology advance for industrial processes. For instance, a similar framework could be used to estimate the interdependencies of structural industrial profiles, in the form of industrial value added, and structural innovative/technological profiles, based on patents. Inference of the underlying network topology could be used to target research and development funds within the firm, and target public investment decisions.

Summing up more generally, the proposed statistical inferential framework may be applied in a variety of productivity network problems, to infer the underlying structure of the network.

The framework developed here is based on complex systems behavior modeling and estimation. There are many possible extensions, left for further studies, including the implementation of out-of-equilibrium time-dependent Ising model.

**Author Contributions:** Author contributions were made in the following areas: conceptualization, C.D., G.R.; methodology, R.F., S.G., M.G.I., L.L.; validation, M.B.B., C.D., S.G., M.G.I., L.L., G.R., W.L.W.; formal analysis, R.F., S.G., M.G.I., L.L., G.R.; investigation, M.B.B., C.D., S.G., W.L.W.; data curation, C.D.; writing—original draft preparation, M.B.B., C.D., R.F., S.G., M.G.I., L.L., G.R., W.L.W.; writing—review and editing, M.B.B., C.D., S.G., W.L.W.; funding acquisition, C.D., M.G.I., L.L., G.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** The financial support of the European Union H2020 Project RISIS 2 (Grant agreement N. 824091) and of Sapienza University of Rome (through the Sapienza Awards no. PH11715C8239C105 and no. RM11916B8853C925) is gratefully acknowledged.

**Acknowledgments:** The data in this paper have been provided by Elsevier within the EBRP Project framework.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

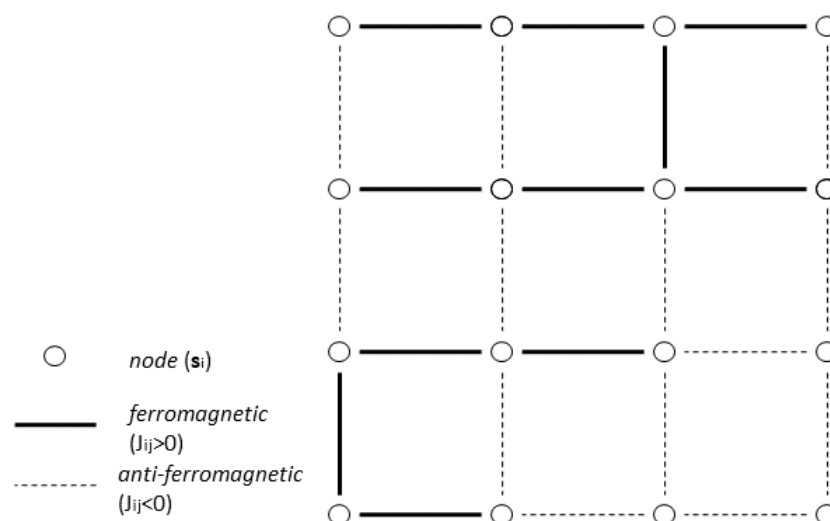
## Abbreviations

The following abbreviations are used in this manuscript:

DEA Data Envelopment Analysis  
NDEA Network Data Envelopment Analysis

## Appendix A. The Ising Spin Glass Model

The Ising spin glass model is made up of a lattice, where each node of the lattice is associated with a vector variable  $\mathbf{s}_i$  at the  $i$ -th site in the  $D$ -dimensional space, that represents the spin of a particle. See Figure A1 for an illustration.



**Figure A1.** Illustration of an Ising Model.  $J_{ij} > 0$  correspond to ferromagnetic couplings;  $J_{ij} < 0$  correspond to anti-ferromagnetic couplings.

Using the relevant Ising and thermodynamic terminology, we can develop a simple spin model to describe the stationary states of our system. For a given couple  $i$  and  $j$  the ‘energy’ unit is given by

$$J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j, \quad (\text{A1})$$



whereas the Hamiltonian is

$$H = -\frac{1}{2}\beta \sum_{ij=1}^N J_{ij} \mathbf{s}_i(t) \cdot \mathbf{s}_j(t) - \sum_{i=1}^N \mathbf{s}_i(t) \cdot \mathbf{h}_i. \quad (\text{A2})$$

The parameters of the system are the intensity of the external magnetic field  $h_i$ , the pairwise interactions  $J_{ij}$ ,  $\beta = (k_B T)^{-1}$ . The Hamiltonian in (A2) describes an Ising model, originally introduced to study the behavior of ferromagnetic systems. The Ising model, when used in the study of different systems, e.g., productivity network, can thus also account for a site-independent weight,  $\beta$ , and external biases,  $h_i$ . For the sake of simplicity we fix here  $\beta = 1$  and  $h_i = 0, \forall i \in (1, N)$ .  $J_{ii} = 0, J_{ij} = J_{ji}$ . The total energy of the system is  $E = \langle H \rangle$ . If the system is in equilibrium at a given 'temperature,'  $T$ , then the energy distribution of the units follows the Boltzmann law, given by

$$F(E) = \frac{1}{Z} e^{-E/k_B T}, \quad (\text{A3})$$

where  $Z$  is the partition function introduced in Section 4.2. The component  $e^{-E/k_B T}$  is known as the Boltzmann factor and  $k_B$  is Boltzmann's constant.

The model assumes ergodicity of the system, i.e., the time average of functions on our spin variables equals the average of these same functions over their probability distributions. This hypothesis is commonly assumed for processes that involve human systems, including time series econometrics.

## Appendix B. Pseudo-Likelihood Approach to Inverse Inference Problem

We report in the following specifics of the objective function used in the optimization algorithm to infer the set of parameters  $\{J\}$ , i.e., the Pseudo-Log-Likelihood function, and of its gradient. The fact that the gradient of the Log-Pseudo-Likelihood function can be calculated exactly makes the computational solution of the inference problem faster and more reliable. By relying on Equations (13) and (14), the expression of the Log-Pseudo-Likelihood function is obtained once one has an explicit expression for the local Hamiltonian,  $H_i(s_i | \{\mathbf{s}_{\setminus i}\}) = -\mathbf{s}_i \cdot [\frac{1}{2} \sum_{j \neq i}^{1,N} J_{ij} \mathbf{s}_j]$  and for the local partition function,  $Z_i = \sum_{\{\mathbf{s}_i\}} e^{-H_i(s_i | \{\mathbf{s}_{\setminus i}\})}$ . We define  $\mathbf{A}_i = \frac{1}{2} \sum_{j \neq i}^{1,N} J_{ij} \mathbf{s}_j$ , thus

$$H_i(s_i | \{\mathbf{s}_{\setminus i}\}) = -\mathbf{s}_i \cdot \mathbf{A}_i \quad (\text{A4})$$

and

$$Z_i = \sum_{\{\mathbf{s}_i\}} e^{\mathbf{s}_i \cdot \mathbf{A}_i} \propto \int_{-1}^1 d\mathbf{s}_i e^{\mathbf{s}_i \cdot \mathbf{A}_i} = \prod_{\gamma=1}^D \int_{-1}^1 ds_i^{(\gamma)} e^{s_i^{(\gamma)} A_i^{(\gamma)}} = \prod_{\gamma=1}^D \frac{2 \sinh(A_i^{(\gamma)})}{A_i^{(\gamma)}}. \quad (\text{A5})$$

The proportionality constant in Equation (A5), equal to the inverse of the total number of all possible  $\mathbf{s}_i$  configurations, does not influence the following derivations and it will be not explicitly considered. The sum  $\sum_{\{\mathbf{s}_i\}} e^{-\mathbf{s}_i \cdot \mathbf{A}_i}$  in Equation (A5) has been calculated by approximating the discrete variables with continuous variables, whose values can continuously vary in the interval  $[-1, 1]$ . The Pseudo-Log-Likelihood function takes thus the expression

$$\lambda(\{J\}) = \sum_{t=1}^T \sum_{i=1}^N [\mathbf{s}_i(t) \cdot \mathbf{A}_i(t) - \sum_{\gamma=1}^D \log(\frac{2 \sinh(A_i^{(\gamma)}(t))}{A_i^{(\gamma)}(t)})] + \text{const.} \quad (\text{A6})$$

The gradient of the Log-Pseudo-Likelihood function is given in Equation (15). To calculate the gradient of the Pseudo-Log-Likelihood with respect to the set of parameters  $J_{ij}$  we thus need to calculate the quantity  $\langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}}$ . It is

$$\begin{aligned} \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}} &= \frac{\sum_{\{\mathbf{s}_i\}} \mathbf{s}_i \cdot \mathbf{s}_j e^{-H_i(s_i|\{\mathbf{s}_i\})}}{\sum_{\{\mathbf{s}_i\}} e^{-H_i(s_i|\{\mathbf{s}_i\})}} = \\ &= \frac{1}{Z_i} \mathbf{s}_j \cdot \int_{-1}^1 d\mathbf{s}_i \mathbf{s}_i e^{\mathbf{s}_i \cdot \mathbf{A}_i} = \frac{1}{Z_i} \sum_{\gamma=1}^D \prod_{\alpha=1}^D s_j^{(\gamma)} \int_{-1}^1 d s_i^{(\alpha)} s_i^{(\gamma)} e^{s_i^{(\alpha)} A_i^{(\alpha)}} = \\ &= \frac{1}{Z_i} \sum_{\gamma=1}^D s_j^{(\gamma)} \left[ \prod_{\alpha \neq \gamma}^{1,D} \frac{2}{A_i^{(\alpha)}} \sinh A_i^{(\alpha)} \right] \frac{2}{[A_i^{(\gamma)}]^2} \left( A_i^{(\gamma)} \cosh A_i^{(\gamma)} - \sinh A_i^{(\gamma)} \right). \end{aligned} \tag{A7}$$

By rephrasing the expression of  $Z_i$  reported in Equation (A5) we obtain

$$Z_i \propto \frac{2 \sinh(A_i^{(\gamma)})}{A_i^{(\gamma)}} \prod_{\alpha \neq \gamma}^{1,D} \frac{2 \sinh(A_i^{(\alpha)})}{A_i^{(\alpha)}}. \text{ Inserting this latter expression in Equation (A7), we get}$$

$$\begin{aligned} \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}} &\propto \sum_{\gamma=1}^D s_j^{(\gamma)} \frac{2}{(A_i^{(\gamma)})^2} \left( A_i^{(\gamma)} \cosh(A_i^{(\gamma)}) - \sinh(A_i^{(\gamma)}) \right) \frac{A_i^{(\gamma)}}{2 \sinh(A_i^{(\gamma)})} = \\ &= \sum_{\gamma=1}^D s_j^{(\gamma)} \left[ \frac{1}{\tanh(A_i^{(\gamma)})} - \frac{1}{A_i^{(\gamma)}} \right], \end{aligned}$$

and finally (the proportionality constant for  $Z_i(\{J\})$  and  $\langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}}$  is the same)

$$\frac{\partial}{\partial J_{ij}} \lambda(\{J\}) = \frac{1}{2} T \left[ Q_{ij} - \frac{1}{T} \sum_{t=1}^T \sum_{\gamma=1}^D s_j^{(\gamma)}(t) \left[ \frac{1}{\tanh(A_i^{(\gamma)}(t))} - \frac{1}{A_i^{(\gamma)}(t)} \right] \right]. \tag{A8}$$

To deal with a lower number of parameters in place of maximizing the Log-Pseudo-Likelihood function, given by the sum of the single-node Log-Pseudo-Likelihood functions (Equation (14)), each single-node Pseudo-Log-Likelihood function is maximized. Since the couplings in the Ising model should be symmetric the final estimate of the  $J_{ij}$  parameter is obtained by taking the average  $(J_{ij} + J_{ji})/2$ .

### Appendix C. Full Scopus 16 Data Summary

**Table A1.** Summary Statistics Performance Variables, by Discipline (ASJC code) (53 Countries, 16 Disciplines, 1996–2012, 901 Observations Each Discipline).

Discipline	Mean	Std. Dev.	Min	Max
AGRI (11)				
Articles	2448.1	4565.2	29	41,261
Highly Cited	328.5	772.9	0	10,068
Number of Authors	4906.4	9635.7	54	97,432
BIOC (13)				
Articles	4357.6	9723.4	44	85,295
Highly Cited	529.1	1572.1	0	15,480
Number of Authors	10,499.4	23,931.7	95	229,139
CENG (15)				
Articles	1211.6	2488.6	11	22,203
Highly Cited	155.0	333.2	0	3557
Number of Authors	2617.4	5788.1	26	65,580

Table A1. Cont.

Discipline	Mean	Std. Dev.	Min	Max
CHEM (16)				
Articles	2686.3	4983.4	31	46,252
Highly Cited	305.0	712.5	0	6460
Number of Authors	5147.5	10,292.8	55	111,789
COMP (17)				
Articles	2727.5	7315.9	7	80,276
Highly Cited	355.6	1013.9	0	10,123
Number of Authors	4161.6	11,548.7	11	128,273
EART (19)				
Articles	1628.3	3360.1	15	24,836
Highly Cited	239.8	582.2	0	4862
Number of Authors	2558.2	5,774.8	20	50,348
ENER (21)				
Articles	635.7	1625.8	3	16,685
Highly Cited	85.0	187.4	0	1974
Number of Authors	1360.0	3771.0	6	47,669
ENGI (22)				
Articles	5433.7	14,078.9	31	156,349
Highly Cited	739.5	1804.5	0	19,830
Number of Authors	8981.8	23,456.8	54	293,605
ENVI (23)				
Articles	1367.7	2827.3	18	22,742
Highly Cited	177.4	400.2	0	3657
Number of Authors	2806.4	6034.3	41	52,584
IMMU (24)				
Articles	1164.6	2406.0	16	18,650
Highly Cited	143.2	395.4	0	3597
Number of Authors	3141.2	6630.4	38	57,748
MATE (25)				
Articles	2751.3	5835.9	22	58,492
Highly Cited	330.5	745.3	0	6939
Number of Authors	4982.2	11,305.5	27	132,822
MATH (26)				
Articles	1804.3	3773.5	7	32,690
Highly Cited	253.9	616.4	0	7565
Number of Authors	2588.5	6174.5	12	63,005
MEDI (27)				
Articles	7672.4	17,175.3	95	165,181
Highly Cited	1,027.9	2929.3	0	28,743
Number of Authors	15,773.2	34,286.3	211	351,702
NEUR (28)				
Articles	968.8	2389.3	2	20,520
Highly Cited	115.8	360.8	0	3340
Number of Authors	2284.5	5636.0	4	53,241
PHAR (30)				
Articles	1103.2	2285.4	12	17,750
Highly Cited	134.6	336.2	0	3385
Number of Authors	2959.7	6299.1	22	53,288
PHYS (31)				
Articles	4396.5	8271.2	19	58,244
Highly Cited	559.9	1208.5	0	10,591
Number of Authors	6946.0	14,391.0	26	127,209

## References

1. Färe, R.; Grosskopf, S. *Intertemporal Production Frontiers: With Dynamic DEA*; Kluwer Academic Publishers: Boston, MA, USA, 1996.
2. Färe, R.; Grosskopf, S. Productivity and intermediate products: A frontier approach. *Econ. Lett.* **1996**, *50*, 65–70. [[CrossRef](#)]
3. Färe, R.; Grosskopf, S. Network DEA. *Socio Econ. Plan. Sci.* **2000**, *34*, 35–49. [[CrossRef](#)]
4. Prieto, A.M.; Zofio, J.L. Network DEA efficiency in input-output models: With an application to OECD countries. *Eur. J. Oper. Res.* **2007**, *178*, 292–304. [[CrossRef](#)]
5. Bogetoft, P.; Färe, R.; Grosskopf, S.; Hayes, K.; Taylor, L. Dynamic network DEA: An illustration. *J. Oper. Res. Soc. Jpn.* **2009**, *52*, 147–162.
6. Fukuyama, H.; Weber, W.L.; Xia, Y. Time substitution and network effects with an application to nanobiotechnology policy for US universities. *Int. J. Manag. Sci.* **2016**, *60*, 34–44. [[CrossRef](#)]
7. Afsharian, M. A frontier-based facility location problem with a centralised view of measuring the performance of the network. *J. Oper. Res. Soc.* **2019**, 1–16. [[CrossRef](#)]
8. Grosskopf, S.; Hayes, K.; Taylor, L.L.; Weber, W.L. Would weighted-student funding enhance intra-district equity in Texas? A simulation using DEA. *J. Oper. Res. Soc.* **2017**, *68*, 377–389. [[CrossRef](#)]
9. Shao, Y.; Bi, G.; Yang, F.; Xia, Q. Resource allocation for branch network system with considering heterogeneity based on DEA method. *Cent. Eur. J. Oper. Res.* **2018**, *26*, 1005–1025. [[CrossRef](#)]
10. Liang, L.; Li, Z.Q.; Cook, W.D.; Zhu, J. Data envelopment analysis efficiency in two-stage networks with feedback. *IIE Trans.* **2011**, *43*, 309–322. [[CrossRef](#)]
11. Chen, Y.; Cook, W.D.; Kao, C.; Zhu, J. Network DEA pitfalls: Divisional efficiency and frontier projection under general network structures. *Eur. J. Oper. Res.* **2013**, *226*, 507–515. [[CrossRef](#)]
12. Chen, Y.; Cook, W.D.; Li, N.; Zhu, J. Additive efficiency decomposition in two-stage DEA. *Eur. J. Oper. Res.* **2009**, *196*, 1170–1176. [[CrossRef](#)]
13. Cook, W.D.; Liang, L.; Zhu, J. Measuring performance of two-stage network structures by DEA: A review and future perspective. *Omega* **2010**, *38*, 423–430. [[CrossRef](#)]
14. Kao, C. Network data envelopment analysis: A review. *Eur. J. Oper. Res.* **2014**, *239*, 1–16. [[CrossRef](#)]
15. Cook, W.D.; Zhu, J. (Eds.) *Data Envelopment Analysis: A Handbook of Modeling Internal Structure and Networks*; Springer: New York, NY, USA, 2014; Volume 208.
16. Kao, C. *Network Data Envelopment Analysis; Foundations and Extensions*; Springer: New York, NY, USA, 2017.
17. Trinh, K.; Zelenyuk, V. *Bootstrap-Based Testing for Network DEA: Some Theory and Applications*; Working Paper No. WP05/2015; School of Economics, University of Queensland: Brisbane, Australia, 2015.
18. Aurell, E.; Ekeberg, M. Inverse Ising inference using all the data. *Phys. Rev. Lett.* **2012**, *108*, 090201. [[CrossRef](#)] [[PubMed](#)]
19. Marruzzo, A.; Tyagi, P.; Antenucci, F.; Pagnani, A. Inverse problem for multi-body interaction of nonlinear waves. *Sci. Rep.* **2017**, *7*, 1–8. [[CrossRef](#)]
20. Ravikumar, P.; Wainwright, M.J.; Lafferty, J.D. High-dimensional Ising model selection using a regularized logistic regression. *Ann. Stat.* **2010**, *38*, 1287–1319. [[CrossRef](#)]
21. Tyagi, P.; Marruzzo, A.; Pagnani, A.; Antenucci, F.; Leuzzi, L. Regularization and decimation pseudo-likelihood approaches to statistical inference in X Y spin models. *Phys. Rev. B* **2016**, *94*, 203–214. [[CrossRef](#)]
22. Georgesçu-Roegen, N. *Entropy Law and the Economic Process*; Cambridge University Press: Cambridge, UK, 1971.
23. Golan, A. Information and entropy econometrics: A review and synthesis. *Found. Trends Econom.* **2008**, *2*, 1–145. [[CrossRef](#)]
24. Golan, A. *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information*; Oxford University Press: Oxford, UK, 2018.
25. Judge, G.G.; Mittelhammer, R.C. *An Information Theoretic Approach to Econometrics*; Cambridge University Press: Cambridge, UK, 2011.
26. Barber, D. *Bayesian Reasoning and Machine Learning*; Cambridge University Press: Cambridge, UK, 2012.
27. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.

28. Koronakos, G.; Chytilova, L.; Sotiros, D. Measuring the Research Performance of UK Computer Science departments via Network DEA. In Proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), Patras, Greece, 15–17 July 2019; pp. 1–7.
29. Lee, B.L.; Worthington, A.C. A network DEA quantity and quality-orientated production model: An application to Australian university research services. *Omega* **2016**, *60*, 26–33. [[CrossRef](#)]
30. Weber, W.L. Network Production and Shadow Prices of Knowledge Outputs. in Rolf Färe, Shawna Grosskopf and Dimitris Margaritis. In *Pricing Non-Marketed Outputs Using Distance Functions*; World Scientific–Now Publishers Series in Business: Hackensack, NJ, USA, 2019; Volume 16, pp. 115–132.
31. Yang, G.; Fukuyama, H.; Song, Y. Measuring the inefficiency of Chinese research universities based on a two-stage network DEA model. *J. Inf.* **2018**, *12*, 10–30. [[CrossRef](#)]
32. Daraio, C. Econometric approaches to the measurement of research productivity. In *Handbook of Science and Technology Indicators*; Glänzel, W., Moed, H.F., Schmoch, H., Thelwall, M., Eds.; Springer: Berlin, Germany, 2019; pp. 633–666.
33. Daraio, C.; Simar, L.; Wilson, P.W. Central limit theorems for conditional efficiency measures and tests of the ‘separability’ condition in nonparametric, two-stage models of production. *Econom. J.* **2018**, *21*, 170–191. [[CrossRef](#)]
34. Färe, R. *Fundamentals of Production Theory*; Springer: Berlin, Germany, 1988.
35. Shephard, R.W. *Theory of Cost and Production Functions*; Princeton University Press: Princeton NJ, USA, 1970.
36. Miller, R.E.; Blair, P.D. *Input-Output Analysis: Foundations and Extensions*; Cambridge University Press: Cambridge, UK, 2009.
37. Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision-making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [[CrossRef](#)]
38. Kemeny, J.; Morgenstern, O.; Thompson, G. A generalization of the von Neumann model of an expanding economy. *Econometrica* **1956**, *24*, 115–135. [[CrossRef](#)]
39. von Neumann, J. Über ein ökonomisches gleichungssystem und eine verallgemeinerung des Brouwerschen Fixpunktsatzes. *Rev. Econ. Stud.* **1945**, *13*, 1–9.
40. Shephard, R.W.; Färe, R. *The Dynamic Theory of Production Correspondences*; Berkeley Operations Research Center, University of California: Berkeley, CA, USA, 1975.
41. Danø, S. *Industrial Production Models: A Theoretical Study*; Springer: New York, NY, USA, 1966.
42. Morroni, M. *Production Process and Technical Change*; Cambridge University Press: Cambridge, UK, 1992.
43. Morroni, M. Production of commodities by means of processes. The flow-fund model, input-output relations and the cognitive aspects of production, *Struct. Chang. Econ. Dyn.* **2014**, *29*, 5–18. [[CrossRef](#)]
44. Morroni, M. *Knowledge, Scale and Transactions in the Theory of the Firm*; Cambridge University Press: Cambridge, UK, 2006.
45. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
46. Sethna, J. *Statistical Mechanics: Entropy, Order Parameters, and Complexity*; Oxford University Press: Oxford, UK, 2006; Volume 14.
47. Hayek, F.A. *Studies in Philosophy, Politics and Economics*; Routledge and Kagan Paul: London, UK, 1967.
48. Hinterberger, F. Self organizing systems. In *The Elgar Companion to Austrian Economics*; Boettke, P.J., Ed.; Edward Elgar: Cheltenham, UK, 1994; pp. 187–191.
49. Sergeev, V.M. The thermodynamic approach to market. *arXiv* **2005**, arXiv:0803.3432.
50. Vozna, L.Y. The notion of entropy in economic analysis: The classical examples and new perspectives. *J. Heterodox Econ.* **2016**, *3*, 1–16
51. Shannon, C.E. A mathematical theory of communication (parts I and II). *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
52. Bostian, M.; Daraio, C.; Grosskopf, S.; Ruocco, G.; Weber, W. Sources and uses of knowledge in a dynamic network technology. *Int. Trans. Oper. Res.* **2020**, *27*, 1821–1844. [[CrossRef](#)]
53. Krapivsky, P.L.; Redner, S.; Ben-Naim, E. *A Kinet. View Stat. Physics*; Cambridge University Press: Cambridge, UK, 2010.
54. Decelle, A.; Ricci-Tersenghi, F. Solving the inverse Ising problem by mean-field methods in a clustered phase space with many states. *Phys. Rev. E* **2016**, *94*, 012112. [[CrossRef](#)] [[PubMed](#)]
55. Nguyen, H.C.; Zecchina, R.; Berg, J. Inverse statistical problems: From the inverse Ising problem to data science. *Adv. Phys.* **2017**, *66*, 197–261. [[CrossRef](#)]

56. Golan, A.; Judge, G.G.; Miller, D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*; Wiley: New York, NY, USA, 1996; pp. 6–38.
57. Brush, S.G. *Science and Culture in the Nineteenth Century: Thermodynamics and History*; University of Texas: Austin, TX, USA, 1967.
58. Besag, J. On the statistical analysis of dirty pictures. *J. R. Stat. Soc. Ser.* **1986**, *48*, 259–302. [[CrossRef](#)]
59. Geman, S.; Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 721–741. [[CrossRef](#)] [[PubMed](#)]
60. Greig, D.M.; Porteous, B.T.; Seheuly, A.H. Exact maximum a posteriori estimation for binary images. *J. R. Stat. Soc.* **1989**, *51*, 271–279. [[CrossRef](#)]
61. Parisi, G. Asymmetric neural networks and the process of learning. *J. Phys. A* **1986**, *19*, L675. [[CrossRef](#)]
62. Banerjee, O.; El Ghaoui, L.; d’Aspremont, A. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **2008**, *9*, 485–516.
63. Bongioanni, I.; Daraio, C.; Ruocco, G. A quantitative measure to compare the disciplinary profiles of research systems and their evolution over time. *J. Inf.* **2014**, *8*, 710–727. [[CrossRef](#)]
64. Hyvarinen, A. Consistency of pseudo-likelihood estimation of fully visible Boltzmann machines. *Neural Comput.* **2006**, *18*, 2283–2292. [[CrossRef](#)]
65. Daraio, C.; Fabbri, F.; Gavazzi, G.; Izzo, M.G.; Leuzzi, L.; Quaglia, G.; Ruocco, G. Assessing the interdependencies between scientific disciplinary profiles at the country level: A pseudo-likelihood approach. *Scientometrics* **2018**, *116*, 1785–1803. [[CrossRef](#)]
66. Nederhof, A.J. The validity and reliability of evaluation of scholarly performance. In *Handbook of Quantitative Studies of Science and Technology*; Elsevier: Amsterdam, The Netherlands, 1988; pp. 193–228.
67. Schubert, A.; Glänzel, W.; Braun, T. Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981–1985. *Scientometrics* **1989**, *16*, 3–478. [[CrossRef](#)]
68. Luwel, M. The use of input data in the performance analysis of R&D systems. In *Handbook of Quantitative Science and Technology Research*; Springer: Dordrecht, The Netherlands, 2004; pp. 315–338.
69. Aksnes, D.; Sivertsen, G.; van Leeuwen, T.N.; Wendt, K.K. Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators. *Sci. Public Policy* **2017**, *44*, 246–258.
70. Moed, H.F. *Citation Analysis in Research Evaluation*; Springer Science & Business Media: Berlin, Germany, 2006.
71. Van Raan, A. Measuring science: Basic principles and application of advanced bibliometrics. In *Handbook of Science and Technology Indicators*; Springer: Cham, Switzerland, 2019; pp. 237–280.
72. Brinn, T.; Jones, M.J.; Pendlebury, M. Measuring Research Quality: Peer Review 1, Citation Indices 0. *Omega* **2000**, *28*, 237–239. [[CrossRef](#)]
73. Madden, G.; Savage, S.; Kemp, S. Measuring public sector efficiency: A study of economics departments at Australian universities. *Educ. Econ.* **1997**, *5*, 153–168. [[CrossRef](#)]
74. Albarrán, P.; Crespo, J.A.; Ortuño, I.; Ruiz-Castillo, J. The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics* **2011**, *88*, 385–397. [[CrossRef](#)]
75. Ruiz-Castillo, J.; Costas, R. The skewness of scientific productivity. *J. Inf.* **2014**, *8*, 917–934. [[CrossRef](#)]
76. Stephan, P.E. *How Economics Shapes Science*; Harvard University Press: Cambridge, MA, USA, 2012.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).