Unfolding the innovation system for the development of countries: co-evolution of Science, Technology and Production

Emanuele Pugliese,[1, 2, 3, *] Giulio Cimini,[4, 1] Aurelio Patelli,[1]
Andrea Zaccaria,[1, 2] Luciano Pietronero,[5, 2] and Andrea Gabrielli[1, 4]

[1]*Istituto dei Sistemi Complessi (ISC)-CNR, 00185 Rome - Italy*
[2]*International Finance Corporation, World Bank Group, 20433 Washington - USA*
[3]*University of Bath, Bath BA27AY - United Kingdom*
[4]*IMT School for Advanced Studies, 55100 Lucca - Italy*
[5]*Sapienza Università di Roma, 00185 Rome - Italy*

We show that the space in which scientific, technological and economic developments interplay with each other can be mathematically shaped using pioneering multilayer network and complexity techniques. We build the tri-layered network of human activities (scientific production, patenting, and industrial production) and study the interactions among them, also taking into account the possible time delays. Within this construction we can identify which capabilities and prerequisites are needed to be competitive in a given activity, and even measure how much time is needed to transform, for instance, the technological know-how into economic wealth and scientific innovation, being able to make predictions with a very long time horizon. Quite unexpectedly, we find empirical evidence that the naive knowledge flow from science, to patents, to products is not supported by data, being instead technology the best predictor for industrial and scientific production for the next decades.

Knowledge production and organization represents the main activity of modern societies – "learning economies" [1] in which most of the wealth of a country is intangible, and the organization of the national innovation system [2], and of diffused creativity [3] are the crucial capabilities for success. Therefore, in the last thirty years the relationships between science, technology and economic competitiveness has become an important focus for social sciences in general and economics in particular [4, 5]. Even though the standard narrative links science, technology and economic productivity in a direct flow [6], actual interactions are likely multi-directional [7] and emerging from the non-trivial interplay among their individual components: the footprint of a complex system.

The new literature of Economic Complexity uses techniques which, differently from traditional social science approach, do not try to average out the complexity of the system, but embraces it by explicitly building on the heterogeneity of individual actors, activities and interactions to extract relevant parameters to characterize the system. Trying to recover the qualitative insights [8] and the few quantitative attempts [9, 10] of the heterodox economists and social scientists, researchers used this approach to study unobservable characteristics and capabilities of countries [11–13], and to unearth unexpected synergies among different activities [14, 15].

Following this line and similarly to [9, 14, 15], here we create the network of interactions between the different human activities. We build on the assumption that if two activities co-occur significantly more often than randomly (in terms of appropriate null models) in the same countries at the same time, then there is an overlap between the capabilities required to achieve proficient level (*i.e.*, competitive advantage) in both. For the first time our network encompasses activities in different realms (or *layers*): scientific fields, technological sectors, and economic production. In such a comprehensive *multi-layer* network [16], *i.e.*, a system where entities belong to different sets and several categories of connections exist among them, we particularly focus on interactions *among* the different layers. As detailed in the Supplementary Information, to task we build the adjacency matrices $M_{c,a}^L(y)$ connecting country c with the activity a belonging to layer L if, in year y, country c was expressing a competitive advantage in activity a with respect to the world average. L stands for the layer of analysis, consisting in the set of all activities related to either Science, Technology or Products export. Note that these layers have an intrinsic hierarchical structure: for instance, in the science layer we can consider activities like *Physics and Astronomy* or corresponding sub-activities (like *Statistical Physics, Condensed Matter Physics, Nuclear and High Energy Physics*). Thus, our matrices do depend on the resolution used for activities classification (even if not explicitly reported in the notation). We use different established databases to construct the multi-layer space: for Science, we take bibliometric data on papers in the period in the various scientific fields from Scopus (`www.scopus.com`); for Technology, we consider the number of patents in different technological sectors extracted from Patstat (`www.epo.org/searching-for-patents/business/patstat`); and for Products export, we use the export data collected by UN COMTRADE (`https://comtrade.un.org/`).

Using these matrices we compute the probability of having a comparative advantage in activity $a_2 \in L_2$ in the year y_2, conditional to having a comparative advantage in activity $a_1 \in L_1$ in the year y_1 (Figure 1), that we define as the

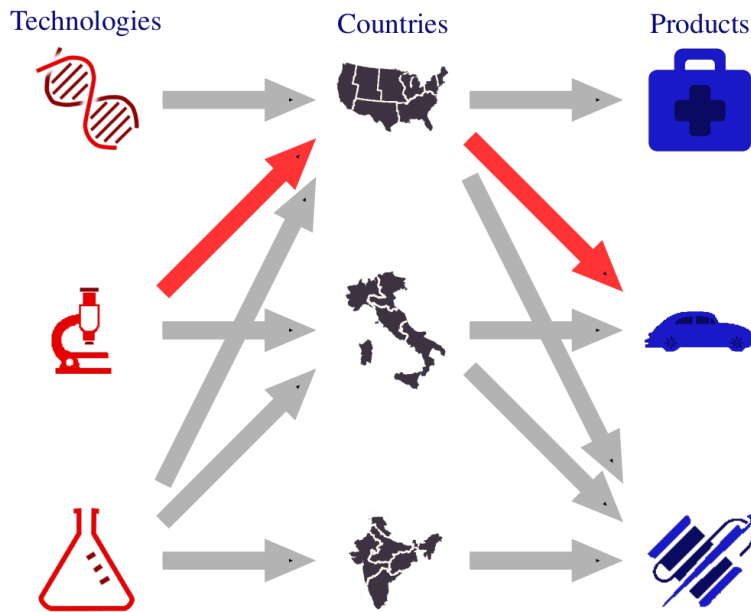arXiv:1707.05146v1 [q-fin.EC] 17 Jul 2017

FIG. 1. Representation of the tri-partite network of *T*echnologies, *C*ountries and *P*roducts. The link direction indicates that we are focusing on the probability of revealing an advantage in a given production conditional to having an advantage in a given technological field. Our measure is also equal to the probability that a bit of information, moving randomly in the directed network, goes from that technology to that product.

Assist matrix, B:

$$B_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2) = Pr(a_2; y_2 | a_1; y_1) = \sum_c Pr(a_2; y_2 | c) Pr(c | a_1; y_1) = \sum_c \frac{M_{c,a_2}^{L_2}(y_2)}{d_c^{L_2}(y_2)} \frac{M_{c,a_1}^{L_1}(y_1)}{u_{a_1}^{L_1}(y_1)},$$

where u_a is the ubiquity of activity $a \in L$, $\sum_c M_{c,a}^L$ and d_c^L is the diversification of country c in the layer L, $\sum_{a \in L} M_{c,a}^L$ (for a similar approach, applied only to the *P*roducts layer, see [15, 17]). Note that we assumed that countries discount all information about capabilities, *i.e.*, that $Pr(a_2; y_2 | c, a_1; y_1) = Pr(a_2; y_2 | c)$. These probabilities can be associated with the overlap between the capabilities required to perform activities a_1 in the year y_1 and a_2 in the year y_2. However, in order to interpret the numbers and assess their significance, we have to compare them with a null model, which we obtain adopting a novel algorithm to randomize bipartite networks [18–20]. Thus, for each pair of (ordered) activities and years, we have an empirical observed value of $B_{a_1 \to a_2}^{L_1, L_2}(y_1, y_2)$ and a distribution of probability for such value under the null hypothesis that the ubiquity of activities and diversification of countries in each layer sum up all the information. This means that activities are independent, and there is no capability structure behind the networks. Therefore co-occurrences between activities happens at random, some more likely than others only because of the ubiquities of the two activities – i.e. some technological fields and the export of some products are less ubiquitous and therefore they are both more likely to be performed by advanced and more diversified countries. Any specific link $a_1 \to a_2$ for which we can reject such null hypothesis is interpreted as the signal of some real interdependency between the capabilities required to a country to perform those specific activities – at the same time or with a time delay.

The described methodology allows obtaining unprecedented qualitative and quantitative insights on the complex dynamics of development. By linking together those activities which are related at a given significance level, we can build the whole multilayer space in which scientific, technological and export activities are embedded (Figure 2). We also perform a more focused analysis and show how a detailed activity (*e.g.*, the export of an individual product) is related to activities in other layers at various aggregation levels. An example is shown in Figure 3, where we plot the scientific and technological fields related to the export of Desktop computers. We can draw from the figure two observations: i) significant peaks, *i.e.*, events with less than 5% probability to occur, are meaningful according to our understanding of the scientific and technological prerequisites to be competitive in Desktop Computer export; ii) technology tend to be more significantly related to export than science, the signal-to-noise ratio being higher. This is not an exception related to this product, as we shall see next.

Looking back at Figure 3, can we say that patenting in the "Electric Digital Data Processing" will lead to better
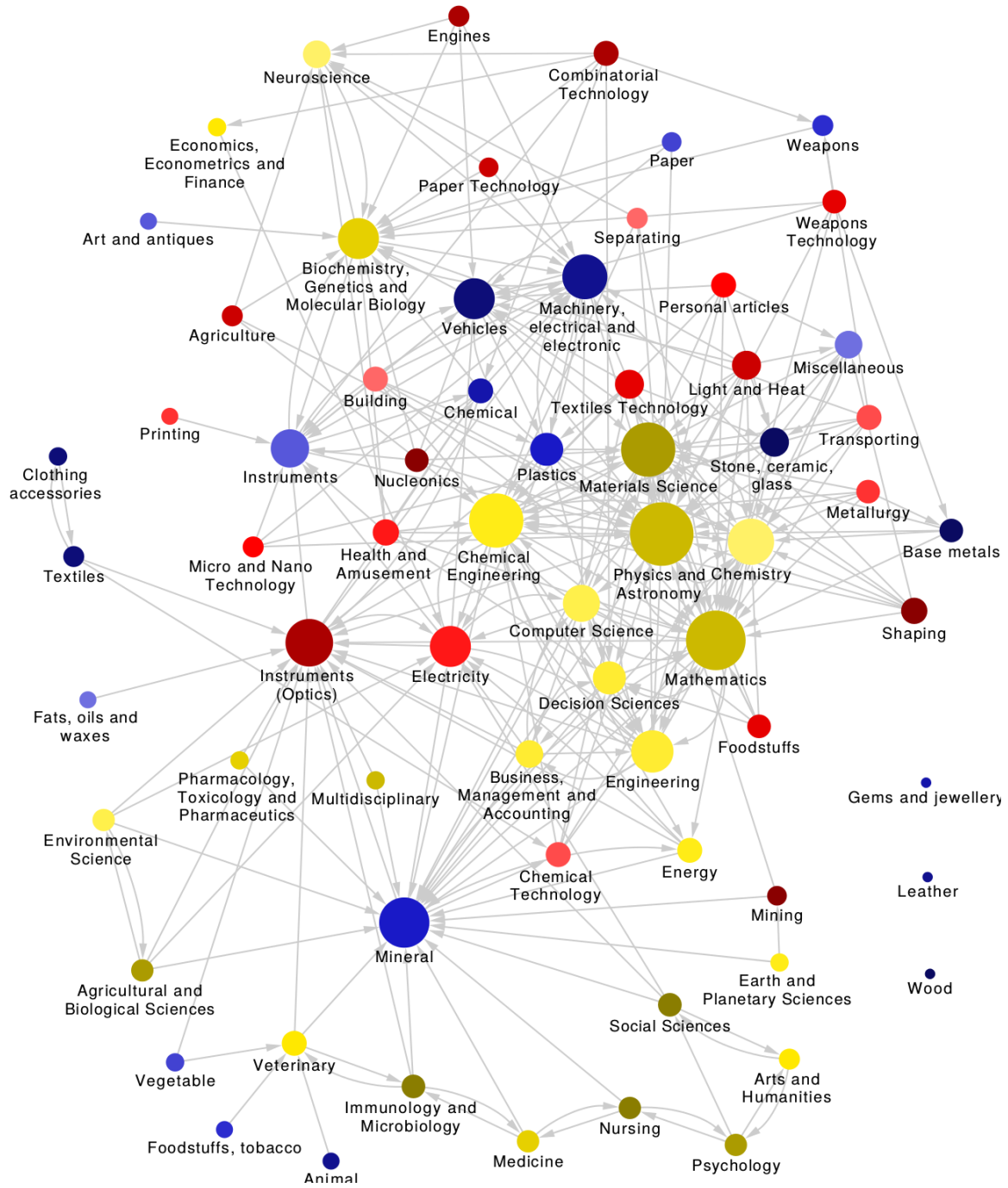
FIG. 2. The multilayer network of broadly aggregated activities (23 scientific major categories, 25 technological sub-sections, 21 product sections). The links are obtained using a significance level of 99.999%. To increase the signal-to-noise ratio, we compute B as the average of three consecutive years in the middle of our sample (2008-2010). Red nodes represent technologies, yellow nodes represent scientific fields and finally blue nodes represent the export of products. The node sizes are proportional to the degree.

performance in the export of Desktop Computers? Since we are handling conditional probabilities, the direct causal interpretation is tempting. However, an inverse causality could be likely as well: countries with a big Desktop Computer industry may be more likely to patent in "Electric Digital Data Processing", or the correlation could arise from the dependence of both variables on a common factor. Despite the several attempts to tackle causality with Big Data analysis [21], the field is far from consensus on these issues. Looking at the time structure in a Granger-non causality fashion [22] can however provide some insights. Given two layers L_1 and L_2 and a time lag Δy we compute
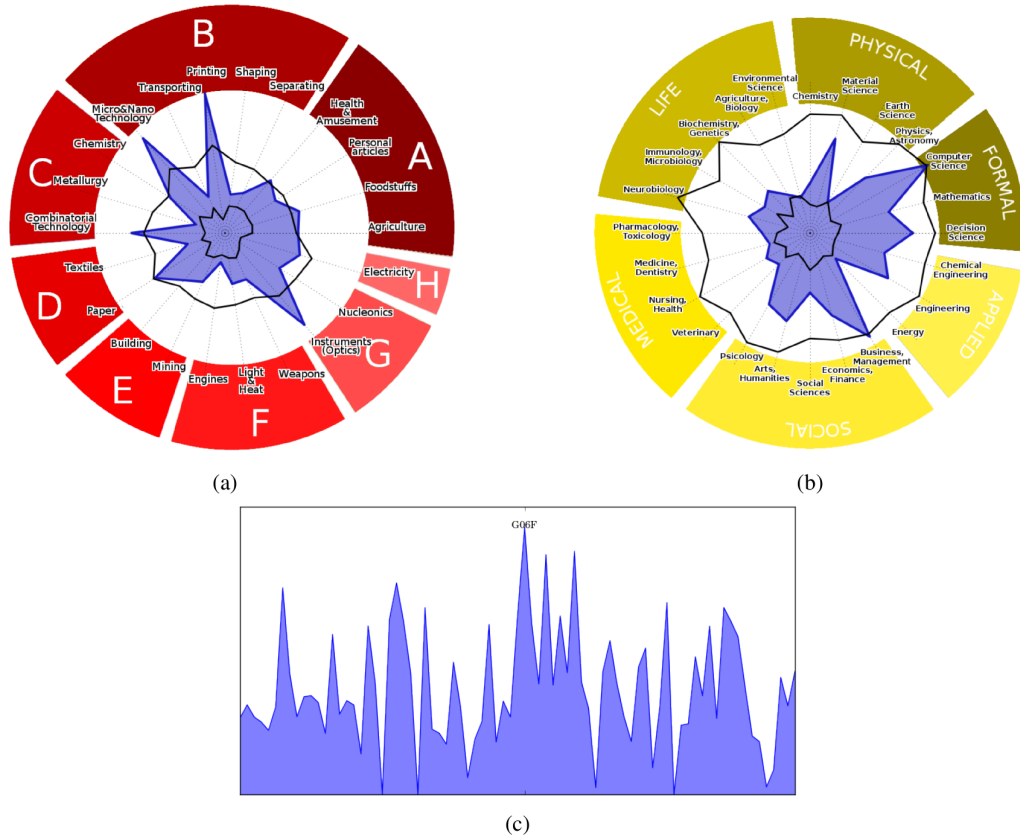
FIG. 3. Technology fields (a) and science fields (b) in 2004-2006 correlated with successful export of Desktop Computers (Harmonized System code 847149) in 2006-2008. The blue contour corresponds to the empirical values of B, the black ones denote the 5% and 95% confidence intervals. (c) Higher resolution analysis with technological section "G: Physics" expanded in its sub-activities. The peak in "G06F" is "Electric Digital Data Processing".

a signal-to-noise ratio $\Phi^{L_1 \to L_2}(\Delta y)$ as the average for different years y of the fraction of significant links $a_1 \to a_2$ in the matrix $B_{a_1 \to a_2}^{L_1 \to L_2}(y, y + \Delta)$. We consider a 95% confidence level, and we consider both links significantly more and significantly less co-occurring than random. Therefore for two unrelated matrices we expect $\Phi \simeq 10\%$ (5% from negative peaks plus 5% from positive ones), and any value above that is the footprint of signal overcoming the noise. Repeating this simple exercise for different temporal windows allows shedding light on the following issue: how many years does it take for the structure of activities in L_1 to influence the activities in L_2?

The results, shown in Figure 4, are striking. First, the signal between two different layers is very high even if far in time, and in both time directions. Beyond the slowly-changing structure of countries activity, this is explained with the strong overlap of countries capabilities on different layers. For instance, the ability to patent successfully in a given technological field is a strong predictor for the successful export of specific products and the publication of papers in specific scientific fields. Second, the technology layer is clearly the best precursor for both science and export, and in both cases the signal reaches its maximum after around 20 years: knowing a country's preferred technologies today gives the highest predictive power for its preferred scientific fields and market sectors in about two decades. However, we are not giving a direct causal interpretation, saying for instance that there is no impact of science on technology. In this specific example the easy explanation is that the chance of someone producing a new patent after reading a scientific paper is not easily appropriable by the country, while possible advances of science due to new technologies (like the discovery of superconductors after advancements in cryogenic techniques) are more likely to be localized in the same country. Moreover, while a technology can have a deep impact on specific scientific fields, scientific research can potentially lead to wide technology spillovers in every fields (think for instance to nuclear physicists at CERN inventing the Internet), which however leave no clear footprint in our signal. Nevertheless we can confidently say that knowing the technological portfolio of a country gives more information on the future scientific fields than the opposite. In the Supplementary Information we apply the same methodology to case studies coming from the other layers.
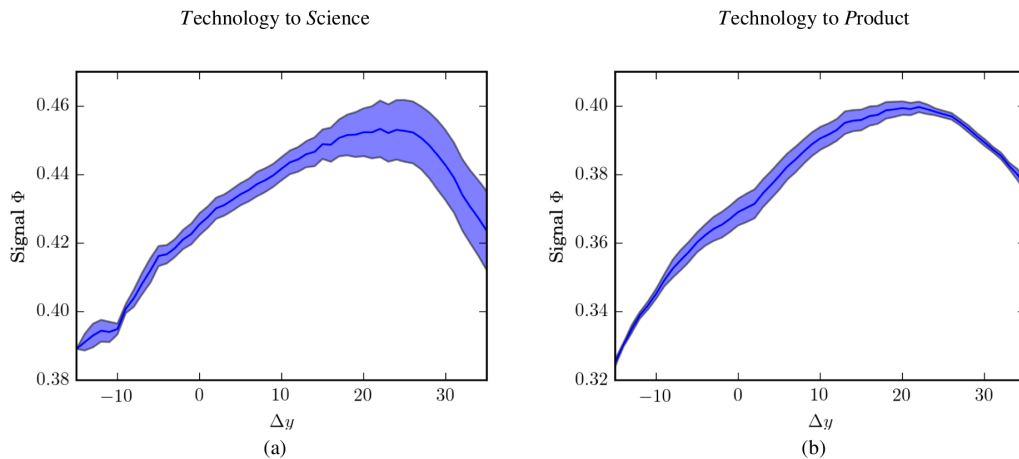
FIG. 4. Signal Φ of *Technology* to *Science* (a) and to *Products* (b) for varying time differences Δy. The time series are build with a sliding window aggregating three years at a time, and looking at all the possible pair of years giving the desired Δy. The blue line and shaded area denote average and standard error of all points corresponding to a given Δy. The analysis is done at an average level of disaggregation: *Technology* is split in subclasses ($\sim$600), *Science* in the categories ($\sim$300) and *Production* at 4 digits level ($\sim$1000). For reference, the noise level is at $\Phi = 10\%$.

Overall, the methods and techniques we presented – in addition to shed light on the dynamics of innovation – can find invaluable use to forecast the scientific fields and market sectors in which countries can have (and will have) a competitive advantage based on their current patent portfolios, even in the long-term: the technical capabilities a country possesses today will define the scientific and market opportunities for the next generation.

SUPPLEMENTARY INFORMATION

DATASET

*S*cience

We use data on scientific productivity and impact of countries collected from the SciVal platform (`www.scival.com`), a new API aggregating data from Scopus (`www.scopus.com`, owned by Elsevier). The database covers journals, trade publications, book series, conference proceedings, and books. Note that while Scopus was shown to have a broader coverage than Web of Science, and to have a more reliable classification system than Google Scholar [23], analyses based on either of these databases usually yields very similar results—especially when performed at the country level [24, 25].

Collected data cover years from 1996 to 2013 and refer to citations accrued by the corpus of scientific publications, each belonging to a given scientific sector (or sub-sector). Data are then aggregated at the country level, so that $W_{c,s}^{S}(y)$ is the number of bibliometric citations obtained overall by the scientific documents produced by country c in scientific sector s during year y. Note that these values are computed using a full counting method for internationally co-authored papers [23], which may cause a bias towards small countries with high level of internationalization to the detriment of large standalone countries [26]. Biases can arise also for anglophone countries in Social Science and Humanities, as documents written in other languages and published in national journals—which are important especially in such sectors—are not covered in full [27, 28]. However, these sectors are supposedly less connected to patenting and productive activities, and as such do not hinder the results of our analyses.

*T*echnology

We use patent data contained in PATSTAT (`www.epo.org/searching-for-patents/business/patstat`) to measure technological capabilities of countries [29]. PATSTAT collects all the patents (approximately 100 millions) by different Patent Offices (almost 100) all around the World. The time span is extremely broad, going from mid-19th

century to today. Patent Offices organize knowledge by tagging each patent with one or more codes from the International Patent Classification (IPC). IPC codes define a hierarchical classification consisting of six levels (*sections, sub-sections, classes, sub-classes, groups, sub-groups*), ranging between 8 and over 70 thousand codes. PATSTAT also records the country of origin of the applicant (usually a firm) of each patent. Finally, PATSTAT defines "families" of patents according to primary citations among them [30], *i.e.*, all patents with common priorities. Indeed, multiple patents could be referred to the same innovation, for example because the same firm applied to different patent offices to extend the protection of their patent to wider markets, or because specific patent offices have heterogeneous regulation about the limits of one patents. In particular we use in the following INPADOC families, or extended family, which collects all the documents that are directly or indirectly linked by one or more priorities.

The basic units of observation are thus the families of patents. Each family is related to one or more countries, through the origin of the applicants of the patents in the family, and one or more technological code. Of course the technological codes depend on the aggregation used. In constructing the matrices, we assume that each family counts as a unit and thus weights accordingly within the matrix. Hence, for each family found in our dataset in a given year, we evenly split its unit of weight among all the technology codes and all the countries it maps to. With this caveats in mind, we define $W_{c,t}^T(y)$ as the number of patent families, or the attributed parts of such families, in the field t applied by a firm located in country c on year y.

Products

To proxy economic production we use the BACI export data, recorded by the UN COMTRADE (https://comtrade.un.org/) and processed by CEPII [31]. The original database reports the import-export flows among countries with a data span 20 years, from 1995 to 2014. This database includes about 5000 products classified according to the Harmonized System 2007 of the World Customs Organization, which denotes them with a set of 6-digits codes organized in a hierarchical way. A given code is divided into three 2-digit parts, each specifying one level of the hierarchy. Hence, the first part indicates the broadest categories, such as "live animals and animal products" (01xxxx) or "plastics and articles thereof" (39xxxx). The second two digits specify further distinctions in each category, such as "live swine" (0103xx) or "live bovines" (0102xx), while the last two digit are even more specific. The trade flows are quantified in thousands of current US dollars. After a data sanitation procedure, a country-product matrix is obtained, whose generic element $W_{c,p}^P(y)$ represents the monetary value of the overall export of country c for product p during year y.

METHODS

Revealed comparative advantage

Given the raw matrices $\{\mathbf{W}^L(y)\}$ for $L \in \{S, T, P\}$ and for the different years y, described in the section above, the first task is to determine whether a given country c shows a comparative advantage in activity a (belonging to layer L), both with respect to other countries as well as to other activities of the same kind. This is achieved through the *revealed comparative advantage* (RCA) [32], an intensive metric computed as the ratio between the weight of activity a in the activity basket of c and the weight of activity a in the total world activity. As a comparative advantage is revealed if $RCA > 1$, we binarize the raw matrices to obtain new matrices $\{\mathbf{M}^L(y)\}$:

$$
M_{c,a}^L(y) = \begin{cases} 1 & \text{if} \quad \dfrac{W_{c,a}^L(y)}{\sum_{a'} W_{c,a'}^L(y)} \Bigg/ \dfrac{\sum_{c'} W_{c',a}^L(y)}{\sum_{c',a'} W_{c',a'}^L(y)} \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}
$$

The use of the RCA-like metrics is common in studies of scientific, technological and economic activities [33–42]. The use of the same RCA formulation for the different layer here is mainly motivated by having a coherent way to build them. Note that the RCA is also properly normalized for the science layer, in order to get rid of the bias towards old papers which had more time to attract citations than recent ones [43].

The multilayer space

Once the binary matrices $\{\mathbf{M}^L(y)\}$ are defined, we build the multi-layer space connecting productive, technological and scientific activities inspired by the general ideas presented in [14, 15]. Given a pair of layers $(L_1, L_2) \in \{P, T, S\}$, in order to assess whether countries having a comparative advantage in activity $a_1 \in L_1$ in year y_1 are more likely to have an advantage also in activity $a_2 \in L_2$ in year y_2, we have to perform an appropriate contraction of $\mathbf{M}^{L_1}(y_1)$ with $\mathbf{M}^{L_2}(y_2)$ over the country dimension (*i.e.*, the set C of countries), and take the element (a_1, a_2). The detailed prescription to build the *Assist* matrix derives from the so called *probabilistic spreading* approach [44].

Let us consider a bit of information located on a generic activity a_1 in layer L_1. We aim at describing how this information can spread to activities in layer L_2. As first step, information jumps to countries according to the connection patterns of $\mathbf{M}^{L_1}(y_1)$: the transition probability that the bit of information goes from a_1 to a given country c is $\rho_{a_1 \to c}^{L_1 \to C}(y_1) = M_{c,a_1}^{L_1}(y_1)/u_{a_1}^{L_1}(y_1)$, where $u_{a_1}^{L_1}(y_1) = \sum_{c \in C} M_{c,a_1}^{L_1}(y_1)$ is the ubiquity (or degree) of a_1 in L_1 for year y_1. We thus assume equal transition probabilities for countries having a comparative advantage in a_1, which is motivated by the maximum uncertainty principle since we do not want to introduce biases in the processes. As second step, information located on countries jumps to activities in layer L_2, now following the connection patters of $\mathbf{M}^{L_2}(y_2)$. Again assuming maximum uncertainty, the transition probability that the bit of information goes from c to a given activity a_2 in layer L_2 is $\rho_{c \to a_2}^{C \to L_2}(y_2) = M_{c,a_2}^{L_2}(y_2)/d_c^{L_2}(y_2)$, where $d_c^{L_2}(y_2) = \sum_a M_{c,a}^{L_2}(y_2)$ is the diversification (or degree) of country c in layer L_2 for year y_2. Putting these two steps together, the probability that the bit of information jumps from activity $a_1 \in L_1$ to activity $a_2 \in L_2$ finally reads:

$$\sum_{c \in C} \rho_{a_1 \to c}^{L_1 \to C}(y_1) \rho_{c \to a_2}^{C \to L_2}(y_2) = \frac{1}{u_{a_1}^{L_1}(y_1)} \sum_{c \in C} \frac{M_{c,a_1}^{L_1}(y_1) M_{c,a_2}^{L_2}(y_2)}{d_c^{L_2}(y_2)} \equiv B_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2). \tag{2}$$

The above equation defines a bipartite network between layers L_1 and L_2, which can be interpreted as the flow of information from activities in L_1 in year y_1 to activities in L_2 in year y_2 (*i.e.*, after a given time). This interpretation is based on the following considerations.

- The bit of information (know-how) associated with a generic activity a_1 is transferred to the various activities in L_2 through the countries having a comparative advantage in a_1.

- In order to account for the highly competitive nature of countries development dynamics (be it scientific, technological or economic), we can naturally assume that transferring the know-how from a_1 to any activity a_2 is not convenient for all countries. In particular, a given country c active in a_1 will put an effort in such a transfer which is inversely proportional to the number of countries having a_1 in their activity baskets. Thus, the probability that country c exploits the information provided by a_1 is given by:

$$Pr(c|a_1; y_1) \equiv \rho_{a_1 \to c}^{L_1 \to C}(y_1). \tag{3}$$

- When a country c transfers its know-how in a_1 to activities in L_2, it is natural to assume that a specific activity a_2 will be chosen with a probability inversely proportional to the number of activities in L_2 in which country c is active. This assumption derives from the finite and fixed amount of resources every country is endowed with for activity transfer. Thus, the conditional probability for the transfer of know-how from activity a_1 to activity a_2 performed by a given country c reads:

$$Pr(a_2; y_2|c; a_1; y_1) = Pr(a_2; y_2|c) \equiv \rho_{c \to a_2}^{C \to L_2}(y_2). \tag{4}$$

The latter equality derives from the assumption that the transfer effort to a_2 is independent from the starting activity a_1 (as well as from y_1). In other words, we model the described random jump as a Markov process: conditional on the present state of the system, its future and past states are independent [45].

Finally, the probability composition formula to assess the transition probability from activity a_1 to activity a_2 leads directly to eq. (2):

$$Pr(a_2; y_2|a_1; y_1) = \sum_{c \in C} Pr(a_2; y_2|c) Pr(c|a_1; y_1) \equiv B_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2). \tag{5}$$

Note that the time direction of the process is given by the time lag $y_2 - y_1$. In the case $y_1 > y_2$, eq. 2 remains unchanged (still representing the transition probability from a_1 to a_2), but the interpretation in terms of information flows is the opposite: $B_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2)$ becomes the probability that a bit of information reaching activity $a_2 \in L_2$ originally came from activity $a_1 \in L_1$.

Hypothesis testing for the multilayer space

To assess the statistical significance of elements of the Assist matrices, we resort to a null model for the bipartite matrices $\{\mathbf{M}^L(y)\}$, built by randomly reshuffling their elements (*i.e.*, the network connections of layer L), but preserving country diversifications and activity ubiquities (*i.e.*, degrees). This allows to wipe out the signal coming from the network connectivity patters, beyond the information contained into the degree values. Yet in order to formalize the null model analytically (and thus avoid relying on a conditional uniform graph test [46, 47]), degree constraints are imposed on average—as for the Canonical ensemble in Statistical Mechanics. We thus end up with a null hypothesis described by the *Bipartite Configuration Model* (BiCM) [18], an extension of the standard *Configuration Model* [48] to bipartite networks.

Formally, the BiCM null model for a given matrix $\mathbf{M}^L(y)$ is defined as the ensemble $\Omega^L(y)$ of bipartite network configurations which are maximally random, except for the ensemble average of the degrees that are constrained to the values observed in the empirical network: $\langle \tilde{d}_c^L(y) \rangle_{\Omega^L(y)} = d_c^L(y) \ \forall c \in C$ and $\langle \tilde{u}_a^L(y) \rangle_{\Omega^L(y)} = u_a^L(y) \ \forall a \in L$. To ease the notation, in the following we omit the explicit dependence of quantities on the layer L and year y, which do not vary throughout the construction of the BiCM. Furthermore, we use symbols with the tilde for quantities assessed on null model configurations, and without the tilde for observed values.

Let $\tilde{\mathbf{M}} \in \Omega$ be a network configuration in the ensemble and $P(\tilde{\mathbf{M}})$ be the probability of that graph within the ensemble. Following the prescriptions from Statistical Mechanics [49], the least biased choice of probability distribution is the one that maximizes the Gibbs entropy

$$S = - \sum_{\tilde{\mathbf{M}} \in \Omega} P(\tilde{\mathbf{M}}) \ln P(\tilde{\mathbf{M}}), \tag{6}$$

subject to the normalization condition $\sum_{\tilde{\mathbf{M}} \in \Omega} P(\tilde{\mathbf{M}}) = 1$ plus the constraints:

$$\langle \tilde{d}_c \rangle_\Omega = \sum_{\tilde{\mathbf{M}} \in \Omega} P(\tilde{\mathbf{M}}) \, \tilde{d}_c(\tilde{\mathbf{M}}) = d_c \quad \forall c \in C, \qquad \langle \tilde{u}_a \rangle_\Omega = \sum_{\tilde{\mathbf{M}} \in \Omega} P(\tilde{\mathbf{M}}) \, \tilde{u}_a(\tilde{\mathbf{M}}) = u_a \quad \forall a \in L. \tag{7}$$

Introducing the respective Lagrange multipliers ω, $\{\mu_c\}_{c \in C}$ and $\{\nu_a\}_{a \in L}$ (one for each country and activity of the network), the probability distribution that maximizes the entropy satisfies, for all configurations $\tilde{\mathbf{M}} \in \Omega$:

$$
\begin{aligned}
0 = \frac{\delta}{\delta P(\tilde{\mathbf{M}})} \Bigg[& S + \omega \left(1 - \sum_{\tilde{\mathbf{M}} \in \Omega} P(\tilde{\mathbf{M}}) \right) + \\
& + \sum_{c \in C} \mu_c \left(d_c - \sum_{\tilde{\mathbf{M}} \in \Omega} P(\tilde{\mathbf{M}}) \, \tilde{d}_c(\tilde{\mathbf{M}}) \right) + \sum_{a \in L} \nu_a \left(u_a - \sum_{\tilde{\mathbf{M}} \in \Omega} P(\tilde{\mathbf{M}}) \, \tilde{u}_a(\tilde{\mathbf{M}}) \right) \Bigg].
\end{aligned}
\tag{8}
$$

The solution is:

$$P(\tilde{\mathbf{M}} \,|\, \{\mu_c\}, \{\nu_c\}) = e^{-H(\tilde{\mathbf{M}} \,|\, \{\mu_c\}, \{\nu_c\})} \Big/ Z(\{\mu_c\}, \{\nu_c\}), \tag{9}$$

where $H(\tilde{\mathbf{M}} \,|\, \{\mu_c\}, \{\nu_c\})$ is the graph Hamiltonian and $Z(\{\mu_c\}, \{\nu_c\})$ is the partition function

$$H(\tilde{\mathbf{M}} \,|\, \{\mu_c\}, \{\nu_c\}) = \sum_{c \in C} \mu_c \, \tilde{d}_c(\tilde{\mathbf{M}}) + \sum_{a \in L} \nu_a \, \tilde{u}_a(\tilde{\mathbf{M}}), \tag{10}$$

$$Z(\{\mu_c\}, \{\nu_c\}) = e^{\omega+1} = \sum_{\tilde{\mathbf{M}} \in \Omega} e^{-H(\tilde{\mathbf{M}} \,|\, \{\mu_c\}, \{\nu_c\})}. \tag{11}$$

Equations (9), (10) and (11) define the BiCM model, namely the distribution over a specified set of network configurations that maximizes the entropy subject to the known constraints. As we are considering local constraints (the degrees), we can work out on eq. (9) to obtain [18]:

$$P(\tilde{\mathbf{M}} \,|\, \{\mu_c\}, \{\nu_c\}) = \prod_{c \in C} \prod_{a \in L} \pi_{c,a}^{\tilde{M}_{c,a}} \, (1 - \pi_{c,a})^{\tilde{M}_{c,a}}, \tag{12}$$

where $\pi_{c,a}$ is the ensemble probability for the connection between country c and activity a:

$$\pi_{c,a} = \langle \tilde{M}_{c,a} \rangle_\Omega = \sum_{\mathbf{M} \in \Omega} \tilde{M}_{c,a} \, P(\tilde{\mathbf{M}} \,|\, \{\mu_c\}, \{\nu_c\}) = \frac{\eta_c \, \theta_a}{1 + \eta_c \, \theta_a} \tag{13}$$

with $\eta_c = e^{-\mu_c}$ and $\theta_a = e^{-\nu_a}$. Note that in eq. (12) the network probability is obtained as the product of connection probabilities over all possible country-activity pairs, meaning that in the BiCM context links results as independent random variables. The probability distribution in eq. (12) yet depends on the values of the Lagrange multipliers, which have to be estimated as:

$$-\frac{\partial}{\partial \mu_c} \ln Z(\{\mu_c\}, \{\nu_c\}) = \langle \tilde{d}_c \rangle_\Omega \quad \forall c \in C, \qquad -\frac{\partial}{\partial \nu_c} \ln Z(\{\mu_c\}, \{\nu_c\}) = \langle \tilde{u}_a \rangle_\Omega \quad \forall a \in L. \tag{14}$$

To obtain the numerical value of the ensemble average of the constraints, we maximize the log-likelihood function:

$$\mathcal{L}(\{\mu_c\}, \{\nu_c\}) = \ln P(\mathbf{M} \,|\, \{\mu_c\}, \{\nu_c\}) = \sum_{c \in C} d_c \ln \eta_c + \sum_{a \in L} u_a \ln \theta_a - \sum_{c \in C} \sum_{a \in L} \ln(1 + \eta_c \, \theta_a), \tag{15}$$

which means solving the system of $|C| + |L|$ equations in $|C| + |L|$ unknowns:

$$\begin{cases} d_c = \sum_{a \in L} \pi_{c,a} = \sum_{a \in L} \dfrac{\eta_c \, \theta_a}{1 + \eta_c \, \theta_a} & \forall c \in C \\[3mm] u_a = \sum_{c \in C} \pi_{c,a} = \sum_{c \in C} \dfrac{\eta_c \, \theta_a}{1 + \eta_c \, \theta_a} & \forall a \in L \end{cases} \tag{16}$$

Connection probabilities of eq. (13) are now well defined, and can be used to directly sample the ensemble of bipartite configurations or to compute the quantities of interest analytically.

Reintroducing in the notation the explicit dependence on the layer L and year y, we finally build the null model $\Omega^{L_1 \to L_2}(y_1, y_2)$ for the Assist matrix from layer L_1 at y_1 to layer L_2 at y_2. This is done by contracting the two BiCMs for the matrices $\mathbf{M}^{L_1}(y_1)$ and $\mathbf{M}^{L_2}(y_2)$ along the country dimension [19, 20], as for eq. (2). We have:

$$\tilde{B}_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2) = \frac{1}{\tilde{u}_{a_1}^{L_1}(y_1)} \sum_{c \in C} \frac{\tilde{M}_{c,a_1}^{L_1}(y_1) \tilde{M}_{c,a_2}^{L_2}(y_2)}{\tilde{d}_c^{L_2}(y_2)}. \tag{17}$$

The distributions of $\tilde{B}_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2)$ values describing the null model can be in principle obtained using numerical [19] or approximated analytical [20] techniques. Due to the non-Gaussianity of such distributions, here we resort to a more practical sampling technique. We thus use eqn. (13) and (17) to generate null Assist matrix networks, and populate the ensemble $\Omega^{L_1 \to L_2}(y_1, y_2)$ to estimate the full distributions. The generic observed element $B_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2)$ is then deemed statistically significant depending on the p-value that we can infer from its distribution under the null hypothesis. The specific threshold for statistical significance and the size of the generated ensemble vary on the exercises performed (as highlighted in the text). It is useful to recall that the two choices, the threshold and the size of the ensemble, are not unrelated: the higher the threshold we want to test, the bigger the sample we require. At the very least, if we set a 99% threshold we naturally need at least 100 realizations of the null model to distinguish empirical points overcoming the threshold, whereas, at least 1000 realizations would be required to test against a 99.9% threshold. While this bare minimum is not enough if we want to test a specific link, it could be enough when, as in Figure 1 of the main text, we just want to check the share of significant links out of a large sample. This is indeed the case of the last exercise, whose results are reported in Figure 4 of the main text. There we fix a pair of layers (L_1, L_2) and a given aggregation level. Then, for a given time lag Δy, we set a measure of signal-to-noise ratio $\Phi^{L_1 \to L_2}(\Delta y)$ equal to average percentage, over the different years y, of significant connections observed in the matrix $\mathbf{B}^{L_1 \to L_2}(y, y + \Delta y)$ using a threshold of 95%. Therefore, we expect $\Phi^{L_1 \to L_2}(\Delta y) \simeq 5\%$ when no signal is found.

Scale of analysis

The scale of the analysis (*i.e.*, the aggregation level of data) is crucial when performing the various exercises. Indeed, even if we do not explicit it in the notation, any analysis can be performed at different aggregation levels, in multiple dimensions. *First*, trivially, we can perform the analysis at different aggregations along the different activities. At a very broad aggregation we can consider "Physics" as one scientific activity, while at a finer aggregation we can consider

any subfield of Physics as an activity. *Second*, we can compute co-occurrences at different geographical aggregations: we can look if two activities co-occur in the same countries, or in the same regions. *Third*, we can look at different temporal aggregations: we can compute the raw matrices $\mathbf{W}^T(y)$ looking at the patents produced in one month, in one year or in five years. The choice of the scale of observation can be a relevant ingredient to look at specific effects: for example the capabilities required at the country level to perform an activity, like "diffused security and education", can be widely different from the capabilities required at the local level, like "infrastructure" or a specific "climatic condition". In other situations, the choice of the scale of observation can be driven mostly by our specific interest in a more or less granular result. This is for example shown in Figure 3 of the main text, where we look at the technological fields required to kick-start the export of "Desktop computers" at different technological resolutions.

There are however practical reasons constraining the possible resolutions which can be accessed. Data availability is indeed a determinant issue. For example, while patents can be assigned to a region at any different scale, both by looking at the address of the inventors and at the address of the assignee firm, exports are recorded by customs and are therefore not easily available at finer geographical resolutions. In general, if we are looking at the same time at scientific papers, patents, and exports, the common geographical resolution cannot be finer than the country, and the time interval cannot be shorter than one year. The second constraining reason is the statistical power of our tests. The finer the activity or time resolution, the less the signal to noise ratio is required to validate each link. This is due to the fact that there will be less activities of a specific kind in a short time, and therefore randomness can play an important role. We have therefore a trade-off, a sort of indetermination principle: if we are interested at specific activities, we have to increase the time window by pooling different years. This can be done both by summing up the matrices for the years in the time interval, or by stacking the matrices by considering different yearly observations of the same country as different rows, as in [15]. This is what we do in the paper when we say that we pool different years. On the contrary, if we are interested at a fine time analysis, like the analysis generating Figure 4 of the main document, either we look at very aggregated fields or, like we do in that analysis, we ignore the specific fields and we simply look at the *number* of significant fields.

———————

   * emanuele.pugliese@gmail.com
 [1] B. Lundvall and B. Johnson, Journal of Industry Studies **1**, 23 (1994).
 [2] R. R. Nelson, *National Innovation Systems: A Comparative Analysis* (Oxford University Press, 1993).
 [3] R. Florida, *The Rise of the Creative Class–Revisited: Revised and Expanded* (Basic Books, 2014).
 [4] P. M. Romer, Journal of Political Economy **98**, S71 (1990).
 [5] G. Dosi, C. Freeman, R. Nelson, G. Silverberg, and L. Soete, eds., *Technical Change and Economic Theory* (Pinter Publishers, 1988).
 [6] B. Bozeman, Research Policy **29**, 627 (2000).
 [7] T. P. Hughes, Social Studies of Science **16**, 281 (1986).
 [8] A. O. Hirschman, "The strategy of economic development," in *Yale studies in economics*, Vol. 10 (Yale University Press, 1958).
 [9] D. J. Teece, R. Rumelt, G. Dosi, and S. Winter, Journal of Economic Behavior and Organization **23**, 1 (1994).
[10] L. Nesta and P. P. Saviotti, The Journal of Industrial Economics **53**, 123 (2005).
[11] C. A. Hidalgo and R. Hausmann, Proceedings of the National Academy of Sciences **106**, 10570 (2009).
[12] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero, Scientific Reports **2**, 723 (2012).
[13] G. Cimini, A. Gabrielli, and F. Sylos Labini, PLoS ONE **9**, e113470 (2014).
[14] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann, Science **317**, 482 (2007).
[15] A. Zaccaria, M. Cristelli, A. Tacchella, and L. Pietronero, PLoS ONE **9**, e113770 (2014).
[16] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, Physics Reports **544**, 1 (2014).
[17] P. Klimek, R. Hausmann, and S. Thurner, PLoS ONE **7**, e38924 (2012).
[18] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini, Scientific Reports **5**, 10595 (2015).
[19] S. Gualdi, G. Cimini, K. Primicerio, R. Di Clemente, and D. Challet, Scientific Reports **6**, 39467 (2016).
[20] F. Saracco, M. J. Straka, R. Di Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini, New Journal of Physics **19**, 053022 (2017).
[21] A. Moneta, D. Entner, P. O. Hoyer, and A. Coad, Oxford Bulletin of Economics and Statistics **75**, 705 (2013).
[22] C. W. Granger, Journal of Economic Dynamics and Control **12**, 551 (1988).
[23] L. Waltman, Journal of Informetrics **10**, 365 (2016).
[24] E. Archambault, D. Campbell, Y. Gingras, and V. Larivière, Journal of the American Society for Information Science and Technology **60**, 1320 (2009).
[25] J. Li, M. Sanderson, P. Willett, M. Norris, and C. Oppenheim, Journal of Informetrics **4**, 554 (2010).
[26] D. W. Aksnes, J. W. Schneider, and M. Gunnarsson, Journal of Informetrics **6**, 36 (2012).

[27] A. J. Nederhof, Scientometrics **66**, 81 (2006).
[28] G. Sivertsen and B. Larsen, Scientometrics **91**, 567 (2012).
[29] "EPO worldwide patent statistical database data catalog," (2014).
[30] "OECD science, technology and industry scoreboard 2001: Towards a knowledge-based economy," (2001).
[31] G. Gaulier and S. Zignago, *BACI: International Trade Database at the Product-Level*, Working Papers 2010-23 (CEPII research center, 2010).
[32] B. Balassa, The Manchester School **33**, 99 (1965).
[33] H. P. Bowen, Weltwirtschaftliches Archiv **119**, 464 (1983).
[34] L. L. Soete, Review of World Economics **117**, 638 (1981).
[35] T. L. Vollrath, Weltwirtschaftliches Archiv **127**, 265 (1991).
[36] D. Archibugi and M. Pianta, Research Policy **21**, 79 (1992).
[37] P. Patel and K. Pavitt, Industrial and Corporate Change **3**, 759 (1994).
[38] S. Kim, The Quarterly Journal of Economics **110**, 881 (1995).
[39] A. Schubert, W. Glänzel, and T. Braun, Scientometrics **16**, 3 (1989).
[40] A. Costinot, D. Donaldson, and I. Komunjer, Review of Economic Studies **79**, 581 (2012).
[41] G. Cimini, A. Zaccaria, and A. Gabrielli, Journal of Informetrics **10**, 200 (2016).
[42] M. R. Guevara, D. Hartmann, M. Aristarán, M. Mendoza, and C. A. Hidalgo, Scientometrics **109**, 1695 (2016).
[43] M. Medo, G. Cimini, and S. Gualdi, Physical Review Letters **107**, 238701 (2011).
[44] T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang, Physical Review E **76**, 046115 (2007).
[45] Y. A. Rozanov, *Markov Random Fields* (Springer Science & Business Media, 2012).
[46] K. Zweig and M. Kaufmann, Social Network Analysis and Mining **1**, 187 (2011).
[47] Z. Neal, Social Networks **39**, 84 (2014).
[48] J. Park and M. E. J. Newman, Physical Review E **70**, 066117 (2004).
[49] K. Huang, *Statistical Mechanics* (Wiley, 1987).