Worldwide bilateral geopolitical interactions network inferred from national disciplinary profiles

M. G. Izzo,[1,2,3,*] C. Daraio,[1] L. Leuzzi,[4,5] G. Quaglia,[1] and G. Ruocco[4,2]

[1]*Universitá degli studi di Roma "La Sapienza," Dipartimento di Ingegneria Informatica Automatica e Gestionale Antonio Ruberti, Via Ariosto, 00185 Roma, Italy*
[2]*Istituto Italiano di Tecnologia - Center for Life Nanoscience, Viale Regina Elena, 291 00161 Roma, Italy*
[3]*Scuola Internazionale Superiore di Studi Avanzati, via Bonomea, 265 34136 Trieste, Italy*
[4]*Universitá degli studi di Roma "La Sapienza," Dipartimento di Fisica, Piazzale Aldo Moro 5, 00185 Roma, Italy*
[5]*Soft and Living Matter Lab, Institute of Nanotechnology, CNR-NANOTEC, Rome, Italy*

A disciplinary profile of a country is defined as the versor whose components are the number of papers produced in a given discipline divided by the overall production of the country. Starting from the Essential Science Indicators (ESI) schema of classification of subject areas, we obtained the yearly disciplinary profiles of a worldwide graph, where on each node sits a country, in the two time intervals 1988–1988 and 1992–2017, the fall of the Berlin Wall being the watershed. We analyze the empirical pairwise cross-correlation matrices of the time series of disciplinary profiles. The contrast with random matrix theory proves that, beyond measurement noise, the empirical cross-correlation matrices bring genuine information. Arising from the Shannon theorem as the least-structured model consistent with the measured pairwise correlations, the stationary probability distribution of disciplinary profiles can be described by a Boltzmann distribution related to a generalized n_d-dimensional Heisenberg model. The set of network interactions of the Heisenberg model has been inferred and to it, two clusterization methods, hierarchical clustering, and principal component analysis have been applied. This allows obtaining a characterization of the worldwide bilateral interactions based on physical modeling. A simple geopolitical analysis reveals the consistency of the results obtained and gives a boost to a deeper historical analysis. In order to obtain the optimal set of pairwise interactions, we used a pseudolikelihood approach. We analytically computed the pseudolikelihood and its gradient. The analytical computations deserve interest in whatever inference Bayesian problem involving an n_d-dimensional Heisenberg model.

I. INTRODUCTION

The key role of high-level specialized knowledge in the nowadays international political relations is an almost undoubted shared opinion. Beyond common sense this idea has been formalized in a sociopolitical context by concepts such as "soft power" [1] or "economic war" [2,3]. Scientific research and technological innovation contribute to the production potential of a State, in turn promoting its economic and, consequently, political power. Losing the production potential can drain it towards an actual political dependence in the worldwide political scenario. The Lisbon European Council held in March 2000 aimed to launch the European Union towards a so-called knowledge-based economy [4], emphasizing how in the contemporary world the culture is somehow thought to be linked to the economy. The strong increase of scientific literature production of emergent countries, such as China or other South-East Asian countries, can be also read under this light. Even if part of the scientific research is actually financed by private companies and most of filed patents belong to them, the role of the State in this dynamics remains preponderant. It retains indeed power in terms of both substantial funding of cutting edge fields and long-term planning of research strategies. The provision of a regulatory environment aimed to support and protect research and innovation, e.g., patent policy, is a further prerogative of the State. Scientific disciplines but not only come into play in these geopolitical dynamics. The legislative system of a country can directly act as an attractiveness of foreign investments, generating national employment and ensuring a tax benefit. The cultural background cast by arts or humanities disciplines contributes to the territorial attractiveness of the country. Economic or computer science can furnish the background for the so-called economic intelligence [5]. Big data analytics, which requires specific informatics, technological, and scientific competencies, can be crucial in the control and creation of information channels. National disciplinary profiles, intended as the high-level and shared knowledge acquired in a given range of specialized fields by a given country, and international relations are thus interconnected since high-level knowledge acquired by the country in a given field can be linked and can promote a specific international political

*mizzo@sissa.it; izzo@diag.uniroma1.it

power. The specialized knowledge, furthermore, supports the flexibility of the country in order to adapt to external changes or to promote them at the expense of less versatile countries. The link between national culture and international relations has been recently the focus of several historical analysis and specific events have been analyzed on this perspective [6–8]. The support of such a sociopolitical hypothesis by quantitative methods has, however, not be achieved. Even the formalization of the matter in a quantitative framework by the definition of proper indices has been questioned [6]. Without going into the merits of the causes, consequences and validity of the political dynamics sketched above, we aim at establishing if this *status quo* can be inferred from a quantitative analysis of the country-level creation of high-level specialized papers. Recently, the quantitative time series analysis and the exploitation of inference methods led to assessing a quantitative network science and machine learning framework in order to study social and economic issues [9]. In Ref. [10], the dynamical structure of political corruption networks has been obtained, the inference method based on the maximum likelihood principle also allowed to obtain predictions for future events. Interestingly, not only network quantitative characterization but even mechanisms inducing the formation of a given network structure can be analyzed [11]. Evolutionary processes of cultural traits have been quantitatively studied in the past [12]. A quantitative and well assessed analysis of their relationship with worldwide geopolitical interactions is, however, lacking. On the other hand, nowadays databases collecting a large amount of specialized manuscripts in different disciplines are available and this permits to obtain measurements of well-defined observables, e.g., bibliometric indicators, in a given interval of time. Since the bibliometric indicators are the *a priori* defined observables, grounding on which a modeling for social analysis can be built, particular care needs to be reserved to their definition and use in the appropriate context [13].

On the basis of the previous observations we assume that (i) it is possible to define a disciplinary profile related to a given country as a vector of n_d elements, being n_d the total number of disciplines taken into account. Each vector component is the relative number of articles related to the corresponding discipline with respect to the total country's production; (ii) in the perspective of analyzing the role of specialized knowledge on the worldwide geopolitical interactions it makes sense to define the disciplinary profile at the country level. These two points permit to identify the national disciplinary profiles as our observables with respect to the matter under exam. Once defined the observables, we quantify their correlations by analyzing their cross-correlation matrices. The comparison with random correlation matrices (RMCs) properties and random matrix theory (RMT) results [14] permits to assess that empirical cross-correlations contain genuine information, related to the characteristics of the underlying network of interactions leading the observed national disciplinary profiles. We further analyze the stationary in time of the genuine information. We then apply maximum-entropy methods derived from the Shannon theorem to model the maximum entropy probability distribution consistent with the measured cross-correlations, without further assumption on nonanalyzed higher-order interactions terms. The resulting

probability distribution is the Boltzmann distribution related to the Hamiltonian of a generalized Heisenberg model with n_d-dimensional spin variables. Development of appropriate tools in the framework of the so-called pseudolikelihood approach, adapted to the generalized Heisenberg model of n_d-dimensional spin variables, permitted us to infer the set of pairwise interactions between different countries. In order to assess the consistency of the inferred interactions and to apply to them clusterization methods, their properties have been contrasted with the ones of RMCs [14–16]. Finally, we apply to the inferred interactions set algorithms usually exploited in the analysis of cross-correlations matrices, hierarchical clustering (HC) and principal component analysis (PCA) [17,18]. We obtain a division in clusters of different countries based on their interactions profile with all the other countries.

We choose to analyze separately two different intervals of time, i.e., 1980–1988 and 1992–2017, the fall of the Berlin Wall being the watershed between the two time intervals. The end of the Cold War marked a rebuilding of the geographic borders. To separate the analysis in the two time intervals is thus unavoidable. A comparison between the interaction profiles inferred in the two time intervals is, however, interesting because the end of the Cold War also marked a change on the nature of international relations. From a geopolitical and military, though not directly acted, plane they moved towards a more distinctly geoeconomic plane. The detailed preliminary analysis of the cross-correlations functions have been performed only for the data referring to the time interval 1992–2017. For sake of simplicity, we analyze only lower order correlations functions, i.e., pairwise. This choice could be *a posteriori* verified if the number of acquisitions would be large enough to obtain a reliable distribution of the N-correlated variables, where N is the total number of variables, as specified in the text. This is not achieved in the present case. The preference of the bilateral interactions with respect to multilateral ones, however, in a geopolitical plane can be identified with a specific tendency of the contemporary world, in particular arising after the end of the Cold War, when the political and economic links of solidarity inside each blocs had not met anymore. The bilateralism of the international relations has been analyzed also in the context of soft power [19,20]. Once consistency of the model and of the inference method has been verified in the statistical groundwork, it is performed a simple sociopolitical analysis. The inferred interactions reflect quite well the international relations drawn by grounding on only geopolitic and geoeconomics arguments.

The paper is organised as follows. Section II contains a definition of the countries disciplinary profiles and describes the database from which the bibliometric data have been extracted and the classification schema. In Sec. III, it is described the analysis of empirical cross correlation matrices. In Sec. IV, it is introduced the inference method, the probability density distribution of the countries disciplinary profiles and the related generalized Heisenberg model with multidimensional spin variables. Furthermore, the analytical computation of pseudolikelihood and its gradient for the generalized Hesenberg model are presented. In Sec. V, the inferred interactions network is characterized applying to it HC and PCA. Section VI contains a simple geopolitical analysis of the

TABLE I. ESI fields used to define the disciplinary profiles.

ESI discipline	Abbreviation
Agricultural Sciences	AGR. SCI.
Biology and Biochemistry	BIO. and BIOC.
Chemistry	CHE.
Clinical Medicine	CL. MED.
Computer Science	COMP. SCI.
Economics and Business	ECO. and BUS.
Engineering	ENG.
Environment/Ecology	ENV. ECO.
Geosciences	GEO.
Immunology	IMMU.
Materials Science	MAT. SCI.
Mathematics	MATH.
Microbiology	MICRO-BIO.
Molecular Biology and Genetics	MOL. BIO. and GEN.
Neuroscience and Behavior	NEUROS. and BEH.
Pharmacology and Toxicology	PHAR. and TOX.
Physics	PHYS.
Plant and Animal Science	PLANT. and AN. SCI.
Psychiatry/Psychology	PSYC.
Social Sciences	SOC. SCI.
Space Science	SPACE SCI.

inferred interactions set. Concluding remarks and outlooks are presented in Sec. VII.

II. DATABASES AND OBSERVABLES

The data analyzed in this paper were extracted from InCites, which is a web-based tool including bibliometric indicators about scientific production and citations of institutions and countries. The indicators are generated from the Web of Science (WoS) documents.[1] The indicators at the country level are created based on address criteria using the whole counting method, i.e., counts are not weighted by numbers of authors or number of addresses, and all addresses attributed to the papers are counted. As subject area scheme for this study, we use the Essential Science Indicators (ESI) schema which comprises 22 subject areas in science and social sciences and is based on journal assignments. Arts and Humanities journals are not included because their coverage, in terms of publication outputs, is lower compared with other subject areas. Each journal is found in only one of the 22 subject areas and there is no overlap between categories. Essential Science Indicators is 22 scientific fields categories in which journals are classified. Only one ESI is assigned to each journal, thus the ESI of a paper will be only one, i.e., the ESI of the journal where it is published. Only papers and reviews from Science Citation Index Expanded and Social Science Citation Index are mapped to ESI. Arts and Humanities, Conference Proceedings Citation Index, and Book Citation Index are excluded. Publications in journals such as

---

[1]The elaborations reported in this paper are based on indicators exported the 2017-09-26 from InCites dataset updated at 2017-09-23 which includes Web of Science content indexed through 2017-07-31.

TABLE II. List of countries over the time period [1980-1988].

COUNTRY	ISO Code
Argentina	ARG
Australia	AUS
Austria	AUT
Belgium	BEL
Brazil	ARG
Bulgaria	BGR
Canada	CAN
Chile	CHL
China	CHN
Czechoslovakia	CZS
Denmark	DNK
Egypt	EGY
Finland	FIN
France	FRA
Germany	DEU
Great Britain	GBR
Greece	GRC
Hungary	HUN
India	IND
Ireland	IRL
Israel	ISR
Italy	ITA
Japan	JPN
Mexico	MEX
Netherlands	NLD
New Zealand	NZL
Nigeria	NGA
Norway	NOR
Poland	POL
South Africa	ZAF
Soviet Union	USSR
Spain	ESP
Sweden	SWE
Switzerland	CHE
Usa	USA
Yugoslavia	SFRJ

Nature or Science, which are multidisciplinary, are assigned by Clarivate Analytics to the most pertinent one using the citations of each publication [21]. The ESI fields used to define the disciplinary profiles are shown in Table I. Over the period 1980–1988, we analyze the scientific production of the 36 countries in Table II and over the period 1992–2017 of the 50 countries reported in Table III. InCites indicators are quite used. Bornmann and Leydesdorf [22], for instance, using InCites indicators, compare normalized citation impact values calculated for China, Japan, France, Germany, United States, and the UK throughout the time period from 1981 to 2010. Since the pioneering works, Refs. [23,24], the characteristics of the disciplinary structure at the country level has been investigated in many studies [25–34]. The analysis of the disciplinary profiles of Eastern Europe contries and Soviet Union and its evolution after the breakup of the Soviet Union has been the subject of several studies [35–39]. Harzing and Giroud [27] comparing the profiles of 34 countries across 21 disciplines showed that nations with the fastest increase in their scientific productivity during the periods 1994–2004 and

TABLE III. List of countries over the time period [1992-2017].

COUNTRY	ISO Code
Argentina	ARG
Australia	AUS
Austria	AUT
Belgium	BEL
Brazil	BRA
Bulgaria	BGR
Canada	CAN
Chile	CHL
China	CHN
Colombia	COL
Czech Republic	CZE
Croatia	HRK
Denmark	DNK
Egypt	EGY
Finland	FIN
France	FRA
Germany	DEU
Great Britain	GBR
Greece	GRC
Hong Kong	HKG
Hungary	HUN
India	IND
Iran	IRN
Ireland	IRL
Israel	ISR
Italy	ITA
Japan	JPN
Malaysia	MYS
Mexico	MEX
Netherlands	NLD
New Zealand	NZL
Norway	NOR
Pakistan	PAK
Poland	POL
Portugal	PRT
Romania	ROU
Russia	RUS
Saudi Arabia	SAU
Singapore	SGP
South Africa	ZAF
South Korea	KOR
Slovenija	SLO
Spain	ESP
Sweden	SWE
Switzerland	CHE
Taiwan	TWN
Thailand	THA
Turkey	TUR
Ukraine	UKR
Usa	USA

2002–2012, which tidied up their disciplinary profile towards a more uniform one, then continued relatively unchanging in their well-proportioned disciplinary structures. Almeida *et al.* [40] and Bongioanni *et al.* [26] examined disciplinary profiles of European countries across 27 disciplines. Thelwall and Levitt [41] analyzed 26 scientific fields in 25 countries and

Pinto and Teixeira [31] examined disciplinary profiles of 65 countries over a broad period of time (1980–2016). Different works analyzed the 16 G7 and BRICS countries [25,39,42,43] exploring the disciplinary profiles of 45 countries. Recently, the disciplinary profiles of countries from all over the world over the years 2009–2019 for the 22 ESI categories from Clarivate Analytics were investigated in Ref. [44].

Our observable is the so-called country disciplinary profile, which is the n_d-dimensional vector,

$$\mathbf{s}_i = (s_i(1), \ldots, s_i(k), \ldots, s_i(n_d)), \tag{1}$$

where i is the index of country, $s_i(k) = n_i(k)/[\sum_k n_i(k)^2]^{1/2}$ being $n_i(k)$ the number of articles created by the country ith in the kth discipline. The vector $\mathbf{s}_i$ has magnitude equal to one, $|\mathbf{s}_i| = 1$. In the following, a μ suffix will denote a given realization of the variable $\mathbf{s}_i$. The index μ in the present case coincides with a temporal index t. Thus $\mathbf{s}_i^\mu$ is a time series related to the ith country. The sampling period of the time series we acquired is one year. The single country can be identified with the node identified by the index i of a network or a graph (world). The number of countries is N, whereas the length of a time series is M. We define furthermore the set of matrix $\mathbf{S} = (S(1), \ldots, S(k), \ldots, S(n_d))$. The matrix $S(k)$ has N rows, the ith row is the time series of the ith country.

III. ANALYSIS OF THE CROSS-CORRELATION MATRICES

The analysis of the spectral measure and of eigenvectors statistics of the empirical cross-correlation functions, together with the contrast with RMT, permits to deal with the following points: (i) to assess if any and identify what features of the measured cross-correlation functions contain genuine information, discerning them from noise content; (ii) to analyze the stationarity in time of the genuine information; (III) to apply under reliable and verified conditions, as stated in the points (i) and (ii) above, maximum-entropy based inference methods under stationarity condition in order to infer the underlying network of interactions generating the observed variables configurations; and (iv) to reduce the number of free parameters in the inference procedure, i.e., to lower the rank of the matrix of interactions, thus possibly allowing inference on shorter time series.

In the following, we will deal only with the first three points above, while leaving the forth point for further developments and only commenting on it in Sec. VII.

As a rationale for putting efforts on the first two points we notice that difficulties on applying inference methods based on maximum entropy models is mostly related to the fact that environment conditions can change in time and the resulting cross-correlations may not be stationary. The finite length of the time series, secondly, introduces measurements noise. If a long time series is used in order to circumvent the problem of finite length, the trouble of nonstationarity of the empirical cross-correlations could in place emerge. The contrast between the spectral measure of the empirical correlations with the one (universal) of RMCs, so-called Wishart matrices, permits to identify the nonrandom components of the measured cross-correlations, which can be thus related to genuine information. As general trend, empirical
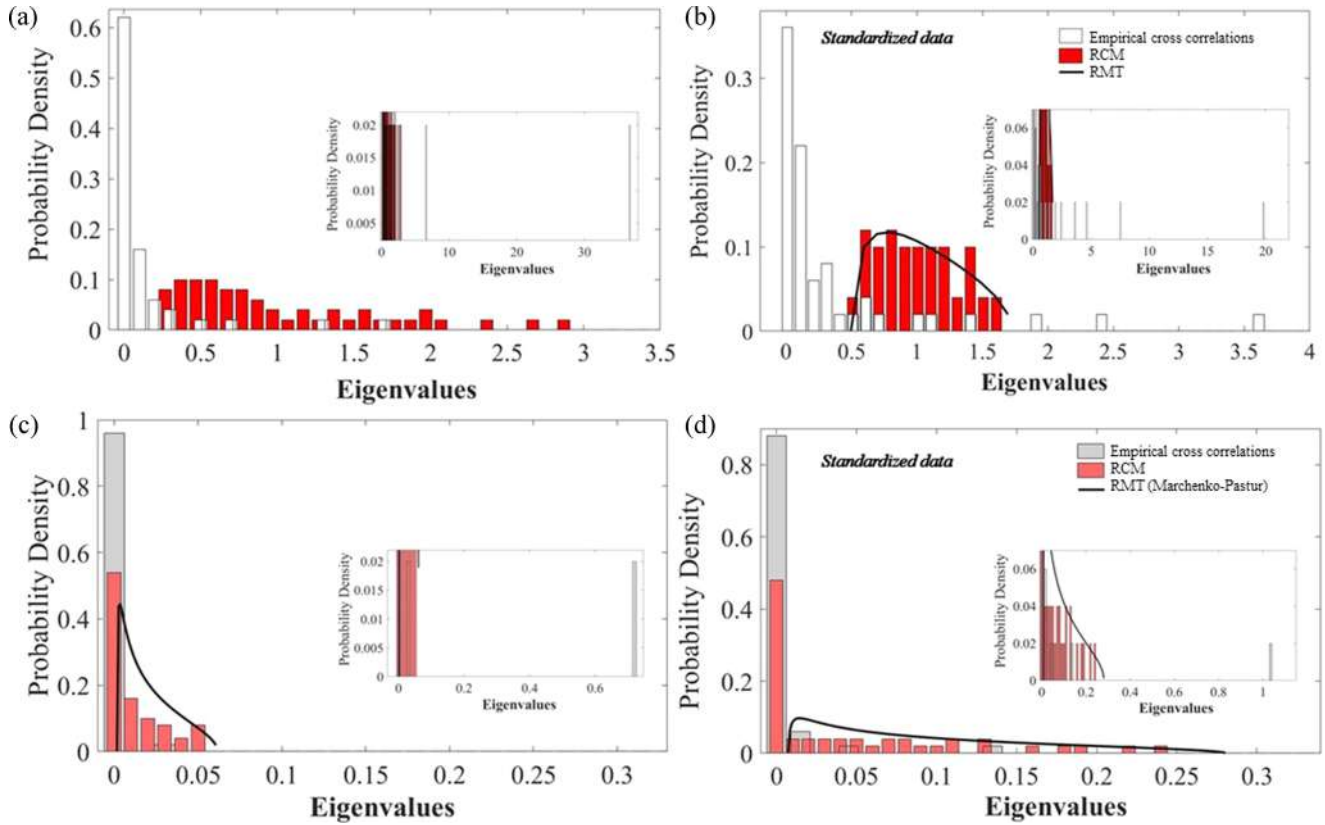
FIG. 1. (a) Histogram representation of the eigenvalues distribution of the empirical correlation matrix $\mathbf{C}$ of disciplinary profiles in the time interval 1992–2017 (white bars) and of the isomorphic finite-dimensional RCM $\mathbf{R}$ (red bars). (b) Histogram representation of eigenvalues distribution of the matrix $\tilde{\mathbf{C}}$ (white bars), obtained from standardized data, and $\tilde{\mathbf{R}}$ (red bars). The solid curve shows theoretical predictions of RMT, obtained from the free convolution of Marchenko-Pastur distributions by R transform. It describes quite accurately the RCM eigenvalues distribution. Discrepancies are observed between the eigenvalues distribution of the empirical cross-correlation matrices and of RCMs in both the low-value and high-value eigenvalues regions (see also the insets). (c) Eigenvalues distribution of $C(k)$ corresponding to MAT. SCI. (light-grey bars) and of $\mathbf{R}(k)$ (light-red bars). The solid curve shows the Marchenko-Pastur distribution. (d) Eigenvalues distribution of $\tilde{\mathbf{C}}(k)$ corresponding to MAT. SCI. (light-grey bars) and of $\tilde{\mathbf{R}}$ (light-red bars). The solid curve shows the Marchenko-Pastur distribution. The insets show a zoom of the large-eigenvalues region of the eigenvalue distribution of empirical covariance matrices.

cross-correlation matrices bringing genuine information are systematically characterized by high-value eigenvalues deviating from the spectral measure of RMCs, which instead matches the so-called bulk-eigenvalue region of the spectral measure [15,16,45]. Furthermore, we will analyze the stability in time of the genuine information contained in the time-dependent cross-correlation matrices, thus allowing the use of results holding for stationary states.

A. Cross-correlations and genuine information

In the following we analyze the properties of the pairwise cross-correlation matrix $\mathbf{C} = \sum_{k=1}^{n_d} \mathbf{C}(k) = \sum_{k=1}^{n_d} \mathbf{S}(k)\mathbf{S}(k)^T$, where the suffix "T" stands for transpose, in the period 1992–2017. The elements of the matrix $\mathbf{C}$ are

$$C_{ij} = \frac{1}{M} \sum_{\mu=1}^{M} \mathbf{s}_i^\mu \cdot \mathbf{s}_j^\mu = \sum_{k=1}^{n_d} C_{ij}(k) = \sum_{k=1}^{n_d} \frac{1}{M} \sum_{\mu=1}^{M} s_i^\mu(k) s_j^\mu(k).$$

(2)

The symbol "·" stands for scalar product. In order to construct RCM isomorphic to the data cross-correlation

matrices we consider a set of n_d independent matrices $\mathbf{\Sigma} = (\mathbf{\Sigma}(1), \dots, \mathbf{\Sigma}(k), \dots, \mathbf{\Sigma}(n_d))$. The matrix $\mathbf{\Sigma}(k)$ is a $N \times M$ matrix of random elements. The mean value and variance of the entries of each matrix $\mathbf{\Sigma}(k)$ is equal to that of $S(k)$. The $\sigma_i^\mu(k)$ entries of the matrix $\mathbf{\Sigma}(k)$, are then properly normalized so that the vector $\boldsymbol{\sigma}_i^\mu = (\sigma_i^\mu(1), \dots, \sigma_i^\mu(k), \dots, \sigma_i^\mu(n_d))$: $|\boldsymbol{\sigma}_i^\mu| = 1$. The RCM $\mathbf{R} = \sum_{k=1}^{n_d} \mathbf{R}(k)$ has elements

$$R_{ij} = \frac{1}{M} \sum_{\mu=1}^{M} \boldsymbol{\sigma}_i^\mu \cdot \boldsymbol{\sigma}_j^\mu = \sum_{k=1}^{n_d} R_{ij}(k) = \sum_{k=1}^{n_d} \frac{1}{M} \sum_{\mu=1}^{M} \sigma_i^\mu(k) \sigma_j^\mu(k),$$

(3)

Figure 1(a) shows the eigenvalues distribution $P(\lambda)$ of the data covariance matrix, $\mathbf{C}$, defined in Eq. (2) contrasted with the eigenvalues distribution of the RCM, $P_{\mathrm{RCM}}(\lambda)$. We also compare the eigenvalues distribution of a single component of $\mathbf{C}$, $\mathbf{C}(k)$, the index k identifying here MAT. SCI. discipline, with the one of a single term of $\mathbf{R}$, $\mathbf{R}(k)$, see Fig. 1(c). The matrix $\mathbf{R}(k)$ is a Wishart matrix $\mathbf{W}$, defined in multivariate statistics [46,47]. The random matrix $\mathbf{R}$ is the sum of Wishart matrices with independent entries, $\mathbf{R} = \sum_{k=1}^{n_d} \mathbf{W}(k)$. Each matrix $\mathbf{W}(k)$ has covariance $\alpha(k)$. The probability

distribution of eigenvalues of a Wishart matrix is asymptotically described, in the limit $N \to \infty$ with $\beta = M/N$ kept constant, by a free Poisson distribution of rate $1/\beta$ and variance $\alpha(k)$. It is known as Marchenko-Pastur distribution,

$$\rho_{\alpha,\beta}(\lambda) = \begin{cases} \tilde{\rho}_{\alpha,\beta}(\lambda), & 0 \leqslant \beta \leqslant 1, \\ (1 - \frac{1}{\beta})\delta_0 + \frac{1}{\beta}\tilde{\rho}_{\alpha,\beta}(\lambda), & \beta > 1, \end{cases} \quad (4)$$

where $\tilde{\rho}_{\alpha,\beta}(\lambda) = \frac{1}{2\pi\alpha\lambda}\sqrt{4\beta\alpha^2 - [\lambda - \alpha(1 + \beta)]^2}$. The symbol δ_0 is the Kronecker delta. We omitted the index k of $\alpha(k)$ in Eq. (4) to make the notation unclaterd. The measure $\tilde{\rho}_{\alpha,\beta}(\lambda)$ is supported on the interval $[\alpha(1 - \sqrt{\beta})^2, \alpha(1 + \sqrt{\beta})^2]$. In Fig. 1(b), it is shown the Marchenko-Pastur distribution (full black line) describing the eigenvalues distribution of $\mathbf{W}(k)$. The random matrix $\mathbf{R}$ is the sum of selfadjoint matrices with spectral measure given by the Marchenko-Pastur distribution. They are characterized by the same parameter β and different variance $\alpha(k)$. The spectral measure of the sum of selfadjoit matrices is given by the free convolution of the spectral measures of each matrix when their size goes to infinity [47,48]. Free convolution can be performed by exploiting the so-called R-transform introduced by Voiculescu [47,49]. In order to gain more insight from RMT predictions we further consider the eigenvalues distribution of crosscorrelation matrix, $\tilde{\mathbf{C}}$, generated by the set of matrices $\tilde{\mathbf{S}} = (\tilde{\mathbf{S}}(1), \dots, \tilde{\mathbf{S}}(k), \dots, \tilde{\mathbf{S}}(n_d))$, as $\mathbf{C}$ is generated by $\mathbf{S}$ following Eq. (2). The entries of the matrices $\tilde{\mathbf{S}}(k)$ satisfy the condition $|\tilde{\mathbf{s}}_i^\mu| = 1$. A further normalization protocol has been applied to $\tilde{\mathbf{S}}(k)$: the average value and variance of its entries is zero and one respectively (standardized data). The eigenvalues distribution of $\tilde{\mathbf{C}}$ is compared to that of a random matrix $\tilde{\mathbf{R}}$ generated by the matrices set $\tilde{\mathbf{\Sigma}} = (\tilde{\mathbf{\Sigma}}(1), \dots, \tilde{\mathbf{\Sigma}}(k), \dots \tilde{\mathbf{\Sigma}}(n_d))$. The entries of each random matrix $\tilde{\mathbf{\Sigma}}(k)$ have zero average and unitary variance, i.e., α is independent from k. The advantage of using the standardized data is that the free convolution of Marchenko-Pastur distributions with same variance α can be easily obtained by using the R transform, differently from when $\alpha = \alpha(k)$. The probability distribution obtained by the free convolution has the same functional form of Eq. (4), but the parameter β is replaced by $n_d\beta$. The asymptotic analytical expression of the spectral measure of $\tilde{\mathbf{R}}$ is thus computed and shown by a full black line in (b) of Fig. 1. In all the analysed cases $P(\lambda)$ shows significant deviations from the corresponding $P_{\text{RCM}}(\lambda)$. In particular we notice a deviating behavior of $P(\lambda)$ of $\mathbf{C}$ and $\tilde{\mathbf{C}}$ in the region of low eigenvalues values, less pronounced in the one-dimensional case [Figs. 1(c) and 1(d)]. Furthermore, it is observed the presence of high-value eigenvalues in the empirical cross-correlation matrices (see the insets in Fig. 1) not reproduced neither by the RMT predictions nor by the spectral measure of the finite-dimensional RCMs.

In order to confirm that the deviating eigenvalues bring genuine information we further analyze the statistics of the corresponding eigenvector components, contrasting it with the one of RCMs and RMT predictions. Figure 2 shows the distribution of components of selected eigenvectors of $\mathbf{C}$ corresponding to (i) its largest eigenvalue max$[\lambda]$, much larger than the largest of the RCM's eigenvalues, λ_+, (ii) a bulk eigenvalue falling inside the RCM band $[\lambda_-, \lambda_+]$, being λ_- the lowest RCM's eigenvalue, and (iii) an eigenvalue lower

than λ_-. Only in the case of the highest eigenvalue significant deviations from RCM's statistics, which is well described by RMT, are observed. Since no information is contained in an eigenvector of a RCM its N-component distribution is a maximum entropy distribution [50], i.e., a Gaussian distribution with zero mean and variance $1/\sqrt{N}$. In Fig. 2(d), it is finally reported the so-called Inverse Participation Ratio (IPR) of eigenvectors of $\mathbf{C}$ and $\mathbf{R}$ as a function of the corresponding eigenvalues. The IPR, I_k, of the eigenvector $\boldsymbol{\xi}_k$ is defined as $I_k = \sum_{i=1}^N \xi_i^{k^4}$. It quantifies the reciprocal number of eigenvector components significantly contributing to it. As it is possible to infer from Fig. 2, the IPR of RCM are localized around an average value, $\langle I \rangle_R = \frac{3}{N}$. The IPR of the eigenvectors of C corresponding to the largest eigenvalue significantly deviate from $\langle I \rangle_R$ and points out a high degree of delocalization of the related eigenvectors. The value of the IPR is indeed well represented by $1/N$ showing that all the components contribute equally. This behavior reveals the delocalized character of the eigenvector, which thus brings information on collective modes of the system [15,16,45,50]. The IPR associated to the lowest eigenvalues also shows deviations from RMT. Since their IPR is larger than $\langle I \rangle_R$ the related eigenvectors are, however, localized on only few nodes. It is worth to observe that the two largest components of eigenvectors corresponding to lowest eigenvalues, $\xi_{\lambda(\text{low})}^i$ and $\xi_{\lambda(\text{low})}^j$, always correspond to large cross-correlation terms C_{ij}.

Finally, we observe that both the single-discipline $\mathbf{C}(k)$ and $\mathbf{R}(k)$ have a finite number of null eigenvalues, see Figs. 1(c) and 1(d). According to their own definition $\mathbf{C}(k)$ and $\mathbf{R}(k)$ have rank $N - M$. The existence of null eigenvalues can hamper the application of mean-field approximation to infere the couplings, since it requires inverting the covariance matrix [51–53]. In parallel, in optimization problems, such as the maximization of the likelihood function described in the following or minimization of the Chi-square function, finite size effects of the time series can make the number of observed configurations much smaller than the number of free parameters and it can lead, e.g., to negative-valued averaged Chi-square. In the vector case, if the components of the vector $\mathbf{s}_i$ are uncorrelated among themselves the number of degree of freedom of the time series of $\mathbf{s}_i$ is increased restoring, e.g., a positive value of the averaged Chi-square. In Ref. [54], we show the eigenvalues of $\mathbf{C}$ and $\mathbf{C}(k)$. The analysis of low-value eigenvalues of the empirical crosscorrelation matrix can also represent a valuable tool to fix the optimal classification scheme of disciplines in the context of our inference problem. If a very dense classification scheme is used, with a very large number of disciplines, correlations among disciplines production will arose introducing redundant information without increasing the degrees of freedom. If a poorly resolved classification scheme is adopted, matters related to small number of degrees of freedom will show up. Correlations among a couple of disciplines is highlighted by considering the number of eigenvalues different from zero of the matrix sum of two single-discipline crosscorrelation matrices calculated for a given time. Dependency between the production of the two disciplines is present if the number of eigenvalues different from zero is one, otherwise
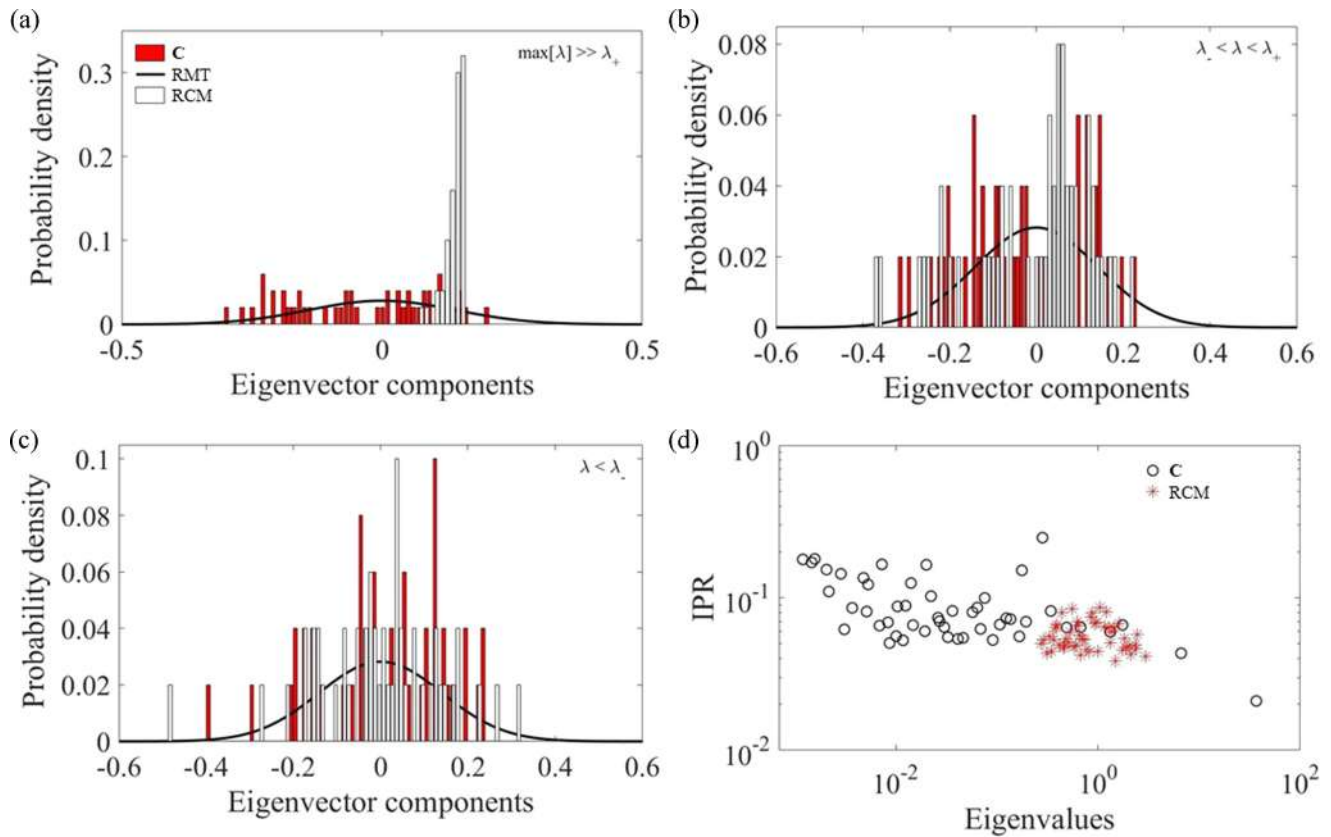
FIG. 2. (a) Distribution of eigenvector components (white bars) corresponding to the largest eigenvalue of **C** ($\lambda \gg \lambda_+$) contrasted with the eigenvector components distribution of test RCM (red bars) and RMT predictions (solid curve). (b) Distribution of eigenvector components corresponding to a bulk eigenvalue λ: $\lambda_- < \lambda < \lambda_+$. (c) Distribution of eigenvector components corresponding to an eigenvalue $\lambda < \lambda_-$. $\lambda_{+(-)}$ is the largest (smallest) eigenvalue of RCM. Significant deviations from RCM's eigenvector components distribution are observed in the case of the eigenvector of **C** corresponding to eigenvalue $\lambda \gg \lambda_+$, in agreement with the findings shown in Fig. 1. (d) Inverse Participation Ratio (IPR) as a function of λ of the empirical covariance matrix (open circles) and of RCM (stars). The IPR of eigenvectors corresponding to both the largest and smallest eigenvalues show deviations with respect to the IPR characteristic of RCM's eigenvectors. The IPR of eigenvectors corresponding to largest and smallest eigenvalues, furthermore, point out a respectively delocalized and localized character.

independency can be assumed. Figure 3 shows at given times for all the possible couple of disciplines the number of eigenvalues different from zero, thus pointing out pair-wise correlations among disciplines.

B. Stationarity properties

We verified that the empirical cross-correlations bring genuine information, pointing out how this information is enclosed in the largest eigenvalues of C and corresponding eigenvectors. We analyze the stability in time of such eigenvectors. Since deviations from RCM outcomes imply genuine correlations *de facto* related to the underlying interactions network, they should show some degree of stability in the time interval used to compute **C** if the interactions network remains stable and if the system is in a stationary state. We define i) the overlap matrix $\mathbf{O}(t, \tau) = \mathbf{V}(t)\mathbf{V}^T(t + \tau)$, where $\mathbf{V}(t)$ is a matrix whose columns are the eigenvectors of the correlataton matrix at time t (notice that for sake of clarity we substitute here the index μ with the index t) corresponding to eigenvalues sorted in ascending order; (ii) the average overlap matrix $\overline{\mathbf{O}}(\tau) = \langle \mathbf{O}(t, \tau) \rangle_t$, i.e., the overlap matrix $\mathbf{O}(t, \tau)$ averaged over all the starting time t included in the measured

time interval. The entries of the matrix $\overline{\mathbf{O}}(\tau)$ for selected values of τ and of $\mathbf{O}(t, \tau)$ for selected values of τ and $t = 1$ are displayed respectively in (a), (b) and (c), (d) of Fig. 4. If all the eigenvectors of the matrix $C(t)$ were nonrandom and stationary, both $\overline{\mathbf{O}}(\tau)$ and $\mathbf{O}(t, \tau)$ would be diagonal with entries equal to one. As shown in Figs. 4 and 5, this condition is approximately satisfied only for eigenvectors corresponding to $\lambda \gg \lambda_+$. In particular, the eigenvector related to the largest eigenvalue remains stable for all the period under consideration, see Fig. 4(d).

IV. INFERENCE METHOD

A. Maximum entropy estimates of second-order marginal and definition of the likelihood function

The Shannon theorem [55] states that the entropy S defined in statistical mechanics is a measure of the "amount of uncertainty" related to a given discrete probability distribution p of variables configuration $\mathbf{s} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$,

$$S[p] = -K \sum_{\{\mathbf{s}\}} p(\{\mathbf{s}\}) \ln[p(\{\mathbf{s}\})]. \tag{5}$$
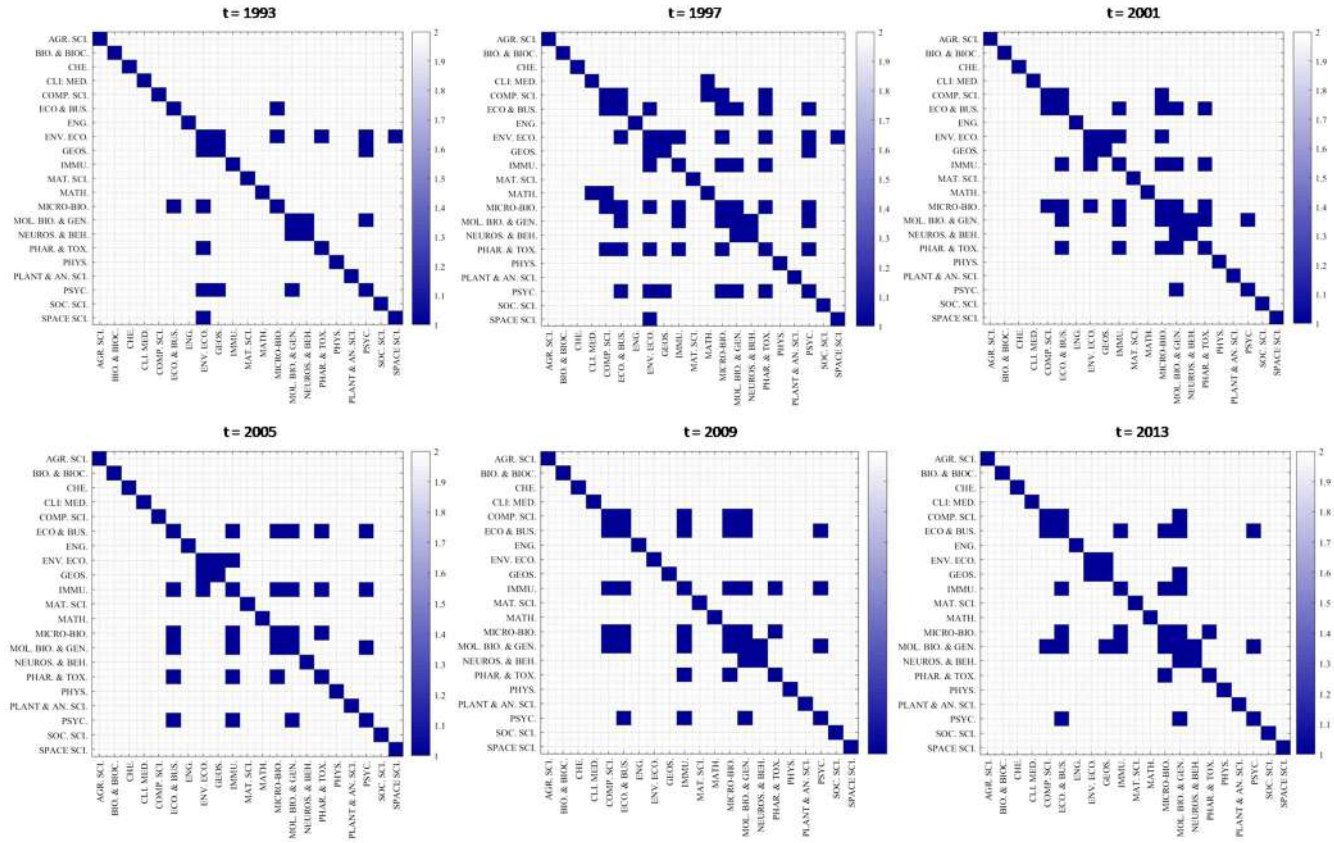
FIG. 3. Number of eigenvalues different from zero in a matrix obtained by the sum of two empirical cross-correlation matrices related to two different disciplines, $\mathbf{C}(k) + \mathbf{C}(l)$. Each cross-correlation matrix is calculated for a single time t. A number of eigenvalues different from zero equal to one highlight a dependency between the manuscripts production of the corresponding couple of disciplines at the time t. Though dependency between some couples of disciplines is present, the production of most of the disciplines is uncorrelated, thus supporting the appropriateness of the classification scheme adopted.

K is a positive constant, hereafter taken equal to one. This quantity is positive and additive for independent sources of uncertainty. In making inference on the basis of partial available information, the probability which maximizes the amount of uncertainty or entropy subject to whatever is known [56–59] has to be used. Since the empirical expectation values are known, formally this means that $p(\{\mathbf{s}\})$ is found as a solution of a constrained optimization problem, i.e., the entropy of the distribution should be maximized subject to conditions that enforce the expectation values to coincide with the empirical ones. One refer to the quantities whose averages are constrained as "features" of the system. As emphasized in Sec. I, we only take into account pairwise interactions, neglecting higher order of interactions. Such a choice could be *a posteriori* validated in the case where an empirical probability distribution of configurations $\{\mathbf{s}\}$ can be obtained from the data, i.e., when the number of acquisitions is large enough [57,58]. This is not the case for the data we are handling. We, however, shortly describe in the following the protocol in order to assess this. The content of information enclosed in a given order interaction (pairwise, triplet, and so on) can be quantified by defining the maximum entropy related to the marginal of order k, where $k = 1, 2, \ldots, N$ being N the total number of observables. Given a joint probability distribution $p(\{\mathbf{s}_1, \ldots, \mathbf{s}_N\})$ the marginal of order k is then defined

as $p_k(\{\mathbf{s}_1, \ldots, \mathbf{s}_k\}) = \sum_{\mathbf{s}_j \neq \mathbf{s}_1, \ldots, \mathbf{s}_k} p(\{\mathbf{s}_1, \ldots, \mathbf{s}_j, \ldots, \mathbf{s}_N\})$. The marginal p_k can be also defined as the maximum entropy distributions that are consistent with the kth order correlations. The marginal of order N corresponds to the exact distribution of the N correlated variables, whereas the marginal of order 1 states for the distribution of N independent variables. The entropies related to marginals of a given order, S_k, decreases monotonically by increasing k towards the true entropy S_N. The connected information or entropy difference, $I_k = S_{k-1} - S_k$, represents the amount by which the maximum possible entropy of the system decreases when one goes from including marginals of order $k - 1$ to including also marginals of order k, thus providing a characterization of the relative importance of various orders of interaction [57,58,60]. The multi-information $I_N = S_1 - S_N$ instead quantify the total amount of correlation in the network, independent of whether it arises from pairwise or higher order interactions. The contrast between I_N and I_2 permits thus to assess if the pairwise interaction model provide an effective description of the system.

Observing variables only in pairs, the optimization problem reduces to

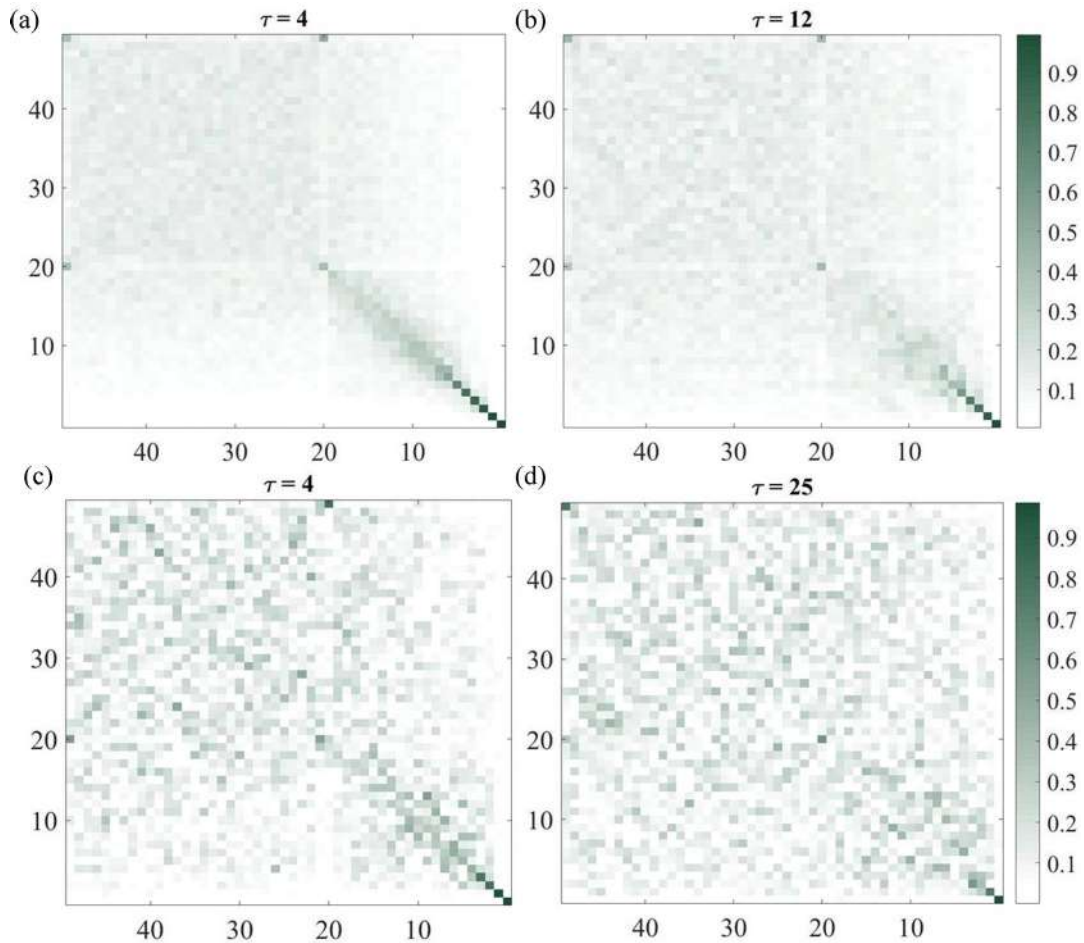$$\text{Max}_{p(\{\mathbf{s}\})} S[p(\{\mathbf{s}\})], \tag{6}$$

FIG. 4. [(a) and (b)] Average overlap matrix, $\overline{\mathbf{O}}(\tau) = \langle \mathbf{O}(t, \tau) \rangle_t$, for two different delay time τ. [(c) and (d)] Overlap matrix $\mathbf{O}(t = 1, \tau)$ for two different delay time τ. The overlap matrix is $\mathbf{O}(t, \tau) = \mathbf{V}(t)\mathbf{V}^T(t + \tau)$, where $\mathbf{V}(t)$ is the matrix whose kth column is the kth eigenvector of the correlation matrix at time t, $\boldsymbol{\xi}_k(t)$, corresponding to the k-th eigenvalue. Eigenvalues are sorted in ascending order. A value of a diagonal entry of the overlap matrix close to one outlines an almost stationarity condition of the corresponding eigenvector. In Sec. III A, it has been shown that eigenvectors corresponding to the largest eigenvalues bring genuine information. Results shown in this figure furthermore emphasize that these eigenvectors are almost stationary.

with the constraints

$$\sum_{\{\mathbf{s}\}} p(\{\mathbf{s}\}) = 1, \qquad (7)$$

$$\langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{p(\{\mathbf{s}\})} = \frac{1}{M} \sum_{\mu=1}^{M} \mathbf{s}_i^\mu \cdot \mathbf{s}_j^\mu. \qquad (8)$$

The features are $f_{ij} = \mathbf{s}_i \cdot \mathbf{s}_j$. The sum is over all possible configurations in the phase space. Equation (6) with the constraints (7) and (8) is solved by using the Lagrange multipliers λ_0, $\{\lambda_{ij}\}$,

$$p(\{\mathbf{s}\}|\{\lambda\}) = \frac{1}{Z} e^{-\frac{1}{2} \sum_{i \neq j} \lambda_{ij} \mathbf{s}_i \cdot \mathbf{s}_j}, \qquad (9)$$

with $Z = \sum_{\{\mathbf{s}\}} e^{-\sum_{i \neq j} \lambda_{ij} \mathbf{s}_i \cdot \mathbf{s}_j}$. The constants $\{\lambda_{ij}\}$ are obtained by the constraint (8). The probability distribution, Eq. (9), is
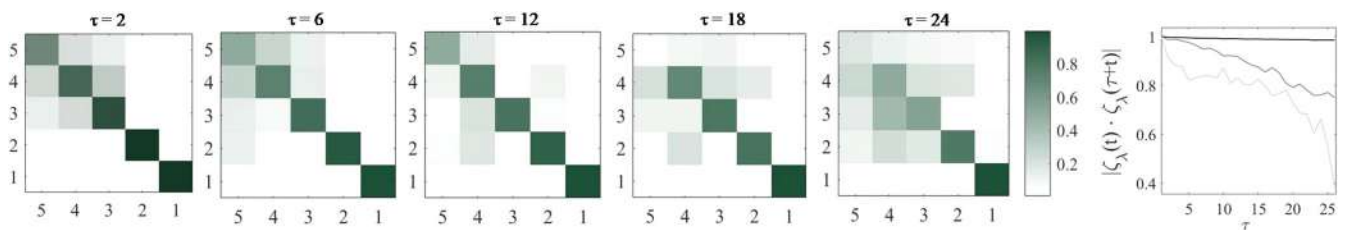

FIG. 5. $O(t = 1, \tau)$ for only the eigenvectors corresponding to the five largest eigenvalues, i.e., the ones bringing genuine information and fulfilling the condition of almost stationarity, at different delay time τ. The autocorrelation, $|\xi_k(t) \cdot \xi_k(t + \tau)|$ of the eigenvectors corresponding to the first three largest eigenvalue is shown for $t = 1$ as a function of τ (right).

a Boltzmann distribution related to a generalized Heisenberg model with pairwise interactions λ_{ij}, n_d-dimensional spin variables and zero external field.

It is possible to reformulate the problem of maximizing the entropy subject to constraint on the expectation values of pairwise correlation functions as searching the maximum of the so-called likelihood function within the class of Boltzmann probability distribution related to the generalized Heisenberg model with multidimensional spin variables. The likelihood function is introduced in the context of Bayesian inference [17]. Given the configuration of a set of spin variables $\{\mathbf{s}\}$ and the data set $\{\mathbf{S}\} = (\{\mathbf{s}^1\}, \ldots, \{\mathbf{s}^\mu\}, \ldots, \{\mathbf{s}^M\})$, it is assumed that i) each realization of $\{\mathbf{s}\}$, $\{\mathbf{s}^\mu\}$, is drawn independently, ii) the data have been generated by a (known) model, which depends on the set of (unknown) pairwise parameters $\{J\}$, with generic element J_{ij}. One aims to find the optimal values of $\{J\}$, which maximize the conditional probability [17]

$$p(\{J\}|\{\mathbf{S}\}) = \frac{p(\{\mathbf{S}\}|\{J\})p(\{J\})}{p(\{\mathbf{S}\})} = \frac{p(\{\mathbf{S}\}|\{J\})p(\{J\})}{\int_{\{J\}} p(\{\mathbf{S}\}|\{J\})p(\{J\})}. \tag{10}$$

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 variable is, as a consequence of Eq. (10), the one that maximizes the likelihood function. By hypothesis the probability distribution of a given configuration $\{\mathbf{s}\}$ belongs to the class of Boltzmann distribution,

$$p(\{\mathbf{s}\}|\{J\}) = \frac{1}{Z(\{J\})} e^{-H(\{\mathbf{s}\}|\{J\})} = \frac{1}{Z(\{J\})} e^{-\frac{1}{2}\sum_{i\neq j}^{1,N} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j}. \tag{11}$$

The set $\{J\}$ with generic elements J_{ij} identify the pairwise interactions of the Heisenberg model. The partition function $Z(\{J\}) = \sum_{\{\mathbf{s}\}} e^{-H(\{\mathbf{s}\}|\{J\})}$. The Hamiltonian or cost function is

$$H(\{\mathbf{s}\}|\{J\}) = \frac{1}{2} \sum_{i\neq j}^{1,N} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j. \tag{12}$$

Given the hypothesis of independence of the data set $\{\mathbf{s}^\mu\}$ and Eq. (11), the log-likelihood function (normalized to the total number of configurations, M), $l(\{J\})$ is given by [17]

$$l(\{J\}) = \ln(L(\{J\})) = \frac{1}{M} \sum_{\mu=1}^{M} -H(\{\mathbf{s}^\mu\}|\{J\}) - \ln(Z(\{J\})). \tag{13}$$

It is immediate to verify, as shown in the following, that the maximum of the log-likelihood function is given by Eq. (9) with λ_{ij} determined by the constraints in Eq. (8), once λ_{ij} have been identified with the pairwise interaction parameters J_{ij}. The gradient of the log-likelihood function is

$$\frac{\partial}{\partial J_{ij}} l(\{J\}) = \frac{1}{2} \left[C_{ij} - \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{\{J\}} \right], \tag{14}$$

where $\langle \ \rangle_{\{J\}}$ states for ensemble average calculated with the probability distribution $p(\{s\}|\{J\})$, Eq. (11), and parameters $\{J\}$. Since, as demonstrated in the following, the log-likelihood is a concave function, the optimal value of J_{ij}, which maximize the log-likelihood, are thus those for which

Eq. (14) is equal to zero, in agreement with the constraint 8 determining $\{\lambda\}$. Under the hypothesis of ergodicity of the system, when the ensemble average is calculated with the "true" set of parameters, i.e., the one whose associated distribution actually generated the data, in the limit $M \to \infty$, $C_{ij} \to \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{\{J\}}$ and $\frac{\partial}{\partial J_{ij}} l(\{J\}) \to 0$. The maximum of $l(\{J\})$ in the limit $M \to \infty$ is thus obtained for those values of $\{J\}$ which generated the correlations C_{ij}. The hypothesis of ergodicity is assumed without further validation.

The Ising or Heisenberg model, have been largely exploited in different fields, beyond the original application to magnets in statistical physics, e.g., in image processing, spatial statistics [61–63] and social networks [64]. It is however worth to observe that by exploiting the Shanon theorem, the Ising or Heisenberg model does not arise from any specific hypotheses about the network but it comes out as the least-structured model consistent with the measured pairwise correlations.

B. Pseudolikelihood approach for the generalized Heisenberg model with n_d-dimensional spin variables

While the definition of the likelihood function has strong theoretical roots, the realization of an optimization algorithm able to draw the optimal $\{J\}$ is hindered by the general intractability of computing the partition function and its gradient. Maximum pseudolikelihood estimation avoids this computational issue entirely by optimizing a different objective function: the pseudolikelihood, which has the advantage to be maximized in polynomial time [65]. The pseudolikelihood function is based on the local conditional likelihood at each node of the network [17,32]. The local conditional probability (single-variable pseudolikelihood) at the ith node is

$$p(\mathbf{s}_i|\{\mathbf{s}_{\backslash i}\}, \{J\}) = \frac{1}{Z_i(\{J\})} e^{-H_i(s_i|\{\mathbf{s}_{\backslash i}\}, \{J\})}, \tag{15}$$

where $\mathbf{s}_{\backslash i}$ indicates the set of all input-variables except the ith. The local hamiltonian $H_i(s_i|\{\mathbf{s}_{\backslash i}\}, \{J\}) = -\mathbf{s}_i \cdot [\frac{1}{2}\sum_{i\neq j}^{1,N} J_{i,j}\mathbf{s}_j]$ and the local partition function is $Z_i(\{J\}) = \sum_{\{\mathbf{s}_i\}} e^{-H_i(s_i|\{\mathbf{s}_{\backslash i}\}, \{J\})}$. By defining $L'(\mathbf{s}_i|\{\mathbf{s}_{\backslash i}\}, \{J\}) = \ln[p(\mathbf{s}_i|\{\mathbf{s}_{\backslash i}\}, \{J\})]$, the normalized log-pseudolikelihood function is

$$l'(\{J\}) = \frac{1}{M} \sum_{\mu=1}^{M} \sum_{i=1}^{N} L'\left(\mathbf{s}_i^\mu \middle| \{\mathbf{s}_{\backslash i}^\mu\}, \{J\}\right). \tag{16}$$

It is possible to show that the pseudolikelihood maximization is exact (i.e., it is maximized by the same set of parameters than the likelihood function) in the limit of infinite sampling [66,67], as discessed in the following. Because in Eq. (12) anti-symmetric piece with respect to the index i and j in the Hamiltonian would cancel, the interactions J_{ij} can be chosen symmetric and the Hamiltonian rephrased as $H = -\sum_{i<j}^{1,N} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j$. The Hessian of both the likelihood and pseudolikelihood functions is thus a triangular matrix. The diagonal elements of the Hessian of the likelihood function, e.g., are $\frac{\partial^2}{\partial J_{ij}^2} l(\{J\}) = \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{\{J\}}^2 - \langle (\mathbf{s}_i \cdot \mathbf{s}_j)^2 \rangle_{\{J\}}$. The latter quantities, apart from some pathological cases where they could be
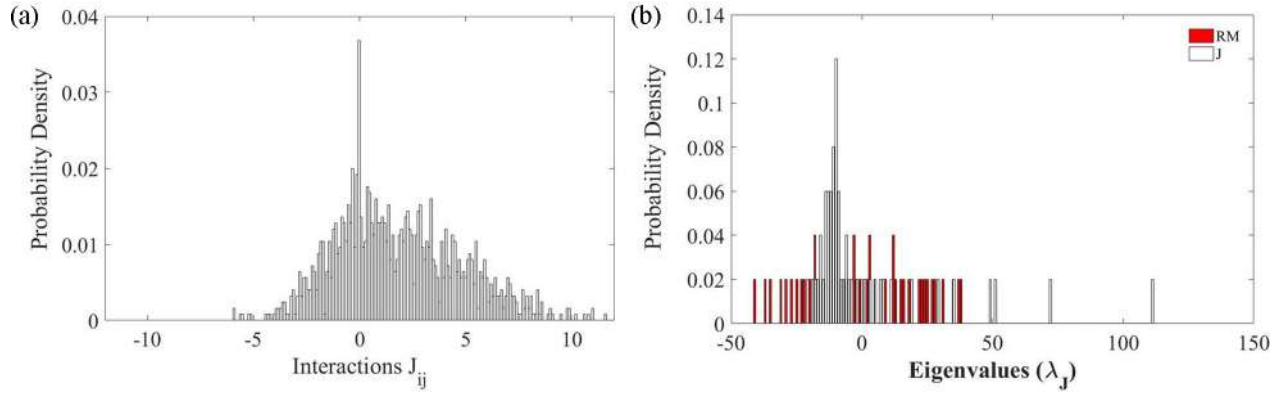
FIG. 6. (a) Histogram representation of the distribution of interactions J_{ij}. (b) Histogram representation of the eigenvalues distribution of the interactions matrix **J** (white bars). The red bars show the eigenvalues distribution of a Gaussian random matrix **R'** isomorphic to **J** (same average and variance of **J** and zero diagonal elements). The nonrandom character of J is emphasized by the discrepancies between the eigenvalue distribution of J and R' observed in the region of high-value eigenvalues, similarly to the case of C.

zero, are negative. Since the eigenvalues of a triangular matrix are the entries on its main diagonal, the likelihood function is strictly concave. Similarly, the pseudolikelihood function is also concave. The gradient of the log-pseudolikelihood function with respect to the parameter J_{ij} is

$$\frac{\partial}{\partial J_{ij}} l'(\{J\}) = \frac{1}{2}\big[C_{ij} - \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}}\big], \tag{17}$$

where $\langle\ \rangle_{i,\{J\}}$ states for ensemble average calculated for the probability distribution $p(\mathbf{s}_i|\{\mathbf{s}_{\setminus i}\}, \{J\})$. It is possible to rephrase the gradient of the log-likelihood function, Eq. (14), obtaining

$$\frac{\partial}{\partial J_{ij}} l(\{J\}) = \frac{1}{2}\big[C_{ij} - \langle \langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}} \rangle_{\{J\}}\big], \tag{18}$$

By comparing Eqs. (17) and (18) it is possible to infer that in the limit $M \to \infty$ (infinite sampling): (i) both the gradients go to zero and (ii) $\frac{\partial}{\partial J_{ij}}\lambda(\{J\}) \to \frac{\partial}{\partial J_{ij}}l(\{J\})$. Because of the concavity of both functions this finally proves the exact maximization of the pseudolikelihood function for $M \to \infty$.

In the case of a n_d-dimensional Heisenberg model with interaction parameters not restricted to nearest neighbor nodes, the partition function and the gradient of the pseudolikelihood function can be calculated analytically, thus facilitating the computational solution of the inference problem through steepest descent method. The partition function is given by

$$Z_i = \int_{-\infty}^{\infty} d\mathbf{s}_i e^{\mathbf{s}_i \cdot \mathbf{A}_i} \delta(s_i - 1), \tag{19}$$

with $\mathbf{A}_i = -\frac{1}{2}\sum_{j=1}^N J_{ij}\mathbf{s}_j$, $s_i = |\mathbf{s}_i|$, and $A_i = |\mathbf{A}_i|$. Introducing polar coordinates (z axis parallel to $\mathbf{A}_i$), so that $\mathbf{s}_i \cdot \mathbf{A}_i =$

$s_i A_i \cos\theta$, it is

$$Z_i = \int_0^{\infty} ds_i\, s_i^{n_d-1} \delta(s_i - 1)\, \omega_{n_d-2} \int_0^{\pi} e^{s_i A_i \cos\theta}(\sin\theta)^{n_d-2} d\theta$$

$$= \omega_{n_d-2} \int_0^{\pi} e^{A_i \cos\theta}(\sin\theta)^{n_d-2} d\theta$$

$$= A_i^{-\frac{n_d-2}{2}}(2\pi)^{\frac{n_d}{2}} J_{\frac{n_d-2}{2}}(iA_i), \tag{20}$$

where ω_n is the area of the unit sphere in n-dimensional space, being $\omega_n = \frac{(2\pi)^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})}$ with $\Gamma(n)$ the Gamma function. The Bessel function of order n, $J_n(t)$ is defined as $J_n(t) = \frac{t^n}{(2\pi)^{n+1}}\omega_{2n}\int_0^{\pi} e^{-it\cos\theta}(\sin\theta)^{2n} d\theta$. To obtain an exact expression of the gradient of the pseudolikelihood function from Eq. (17) it is needed to calculate $\langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}}$. We find

$$\langle \mathbf{s}_i \cdot \mathbf{s}_j \rangle_{i,\{J\}} = \frac{1}{Z_i} \int_{-\infty}^{\infty} d\mathbf{s}_i e^{\mathbf{s}_i \cdot \mathbf{A}_i} \mathbf{s}_i \cdot \mathbf{s}_j \delta(s_i - 1)$$

$$= \frac{1}{Z_i}\sum_{\alpha=1}^{n_d} \frac{\partial}{\partial A_i^{\alpha}} Z_i s_j^{\alpha} = \frac{J_{\frac{n_d+1}{2}}(iA_i)}{J_{\frac{n_d-2}{2}}(iA_i)}\hat{\mathbf{A}}_i \cdot \mathbf{s}_j, \tag{21}$$

where $\hat{\mathbf{A}}_i = \frac{\mathbf{A}_i}{A_i}$. In the case that the dimension n_d is odd the Bessel function can be analytically expressed in terms of elementary functions, obtaining

$$Z_i = \omega_{n_d-2} 2 \sum_{k=0}^{\nu}\Bigg[\frac{\nu!}{k!(\nu-k)!}(-1)^k$$

$$\times \sum_{l=0}^{k}\frac{(2k)!}{(2k-2l)!}\frac{1}{A_i^{2l}}\left(\frac{\sinh A_i}{A_i} - \frac{\cosh A_i}{A_i^2}(2k-2l)\right)\Bigg], \tag{22}$$

where $\nu \in N : n_d - 2 = 2\nu + 1$. Furthermore

$$\frac{\partial}{\partial A_i^{\alpha}} Z_i = \omega_{n_d-2} 2 \sum_{k=0}^{\nu}\frac{\nu!}{k!(\nu-k)!}(-1)^k \sum_{l=0}^{k}\frac{(2k)!}{(2k-l)!}\frac{A_i^{\alpha}}{A_i^{2l+1}}\left(\sinh A_i - \frac{\cosh A_i}{A_i}(2k-l-1)\right.$$

$$\left. - \frac{\sinh A_i}{A_i^2}(2k-l-1) + 2\frac{\cosh A_i}{A_i^3}(2k-l)\right). \tag{23}$$
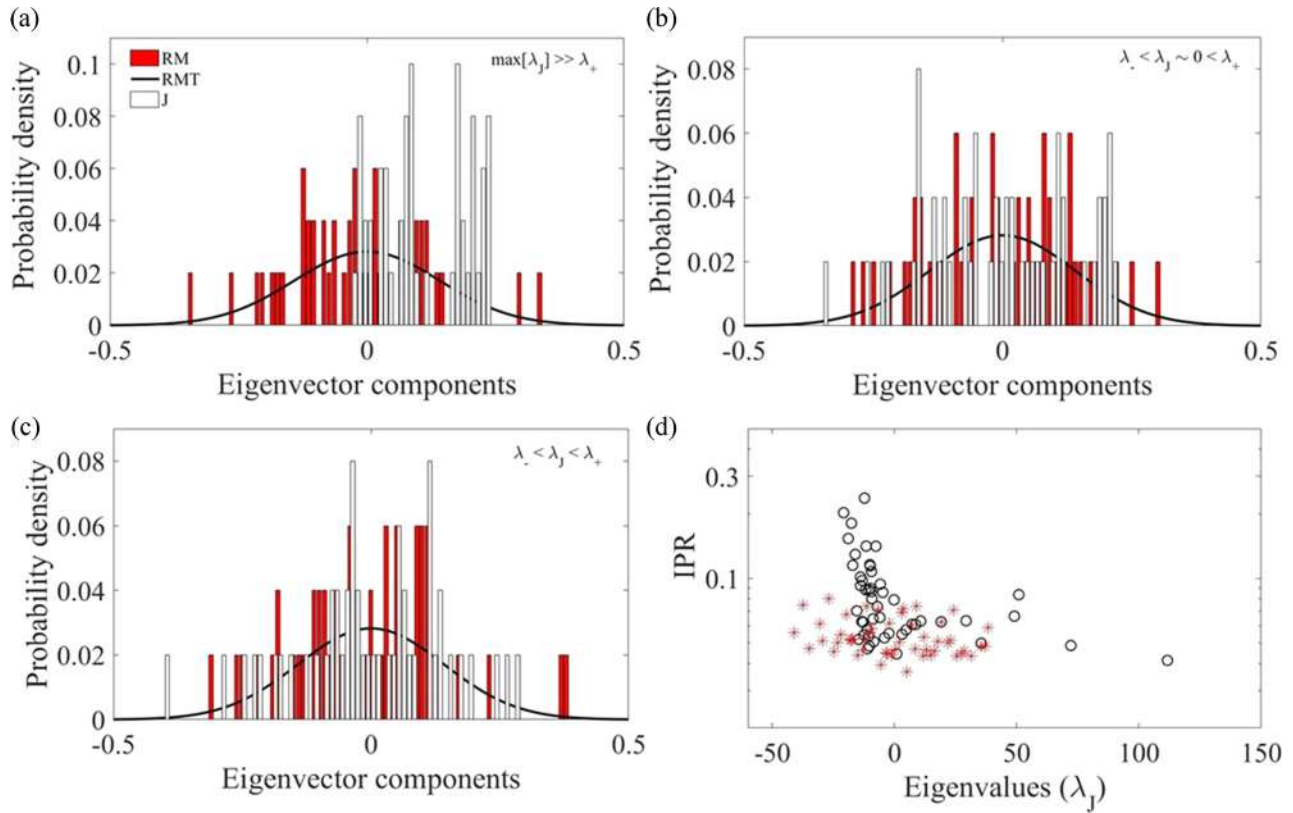
FIG. 7. (a) Distribution of eigenvector components (white bars) corresponding to the largest eigenvalue of the interactions matrix **J**, $\lambda_J \gg \lambda_+$, contrasted with the eigenvector distribution of a test Gaussian random matrix (red bars) and RMT prediction (solid line). (b) Distribution of eigenvector components corresponding to a bulk eigenvalue λ_J: $\lambda_- < \lambda_J < \lambda_+$ and $\lambda_J \sim 0$. (c) Distribution of eigenvector components corresponding to an eigenvalue λ_J: $\lambda_- < \lambda_J < \lambda_+$. $\lambda_{+(-)}$ is the largest (smallest) eigenvalue of the test Gaussian random matrix. As in the case of cross-correlation matrix, significant deviations from random matrix eigenvector components distribution are observed in the case of the eigenvector of **J** corresponding to eigenvalue $\lambda_J \gg \lambda_+$, in agreement with the findings shown in Fig. 6. (d) Inverse Participation Ratio (IPR) as a function of λ_J of **J** (open circles) and a test Gaussian random matrix (stars). The largest and smallest eigenvalues, as in the case of **C** have an IPR deviating from the one of the Gaussian random matrix, showing respectively a delocalized and localized character.

The maximization of the log-log-pseudolikelihood functions has been performed by means of the MATLAB FMINUNC package [68] by selecting a trust-region optimization algorithm. A *l2*-regularizer (parameter 0.13) was used [69,70].

V. THE INFERRED INTERACTIONS NETWORK

A. Assessment of the inferred interactions J

In order to attest the consistency of the inference protocol as well as to analyze the content of information contained in the set of inferred interactions, the spectral measure and the eigenvectors of **J** are contrasted with corresponding RCMs. Figure 6(a) shows a histogram representation of the probability distribution of the values of J_{ij}. We define a random matrix **R**', whose entries are extracted by a normal distribution. Its mean value and variance are set to the ones of $\{J\}$, the diagonal elements are set to zero and the matrix is furthermore made symmetric. The spectral measure of **J**, displayed in (b) of Fig. 6, is compared to the one of **R**'. Similarly to the case of the empirical cross-correlations matrices, the spectral distribution of **J** covers larger values than the one of **R**'. The comparison between the probability distribution of the eigenvectors (ζ_k) components related to selected eigenvalues (λ_J) of **J** and **R**' confirms the nonrandom character of the eigen-

vector associated to the largest eigenvalue, see Figs. 7(a)–7(c). The IPR of the eigenvectors of **J** and **R**', displayed in Fig. 7(d) as a function of λ_J, further reveals a deviation from RMT results for largest and smallest eigenvalues. As in the case of the cross-correlations, the eigenvectors corresponding to largest eigenvalues have a delocalized character, whereas the ones associated to lowest eigenvalues show strong localization. The eigenvector corresponding to the largest eigenvalues are thus related to the whole structure of the interactions network, so in the case of empirical cross-correlations the eigenvector related to largest eigenvalue carry information about collective modes of the system. Similarly to the case of the empirical cross-correlations, the eigenvectors associated to the lowest eigenvalues are sensitive to the largest values of J_{ij}'s. Figure 8 shows the eigenvector composition of J (full lines) and C (dashed lines) corresponding to the respective first two lowest eigenvalues. The first two largest components of the eigenvector related to the lowest λ_J (full black line) correspond to Russia and Ukraine and the largest interaction J_{ij} is the one between the same two countries. It exists a correlation between the values of C and of J, as it is possible to infer by observing Fig. 9(a), which shows the values of J_{ij} as a function of the values of C_{ij}. The nature of such a correlation can be better understood when the absolute value of the scalar
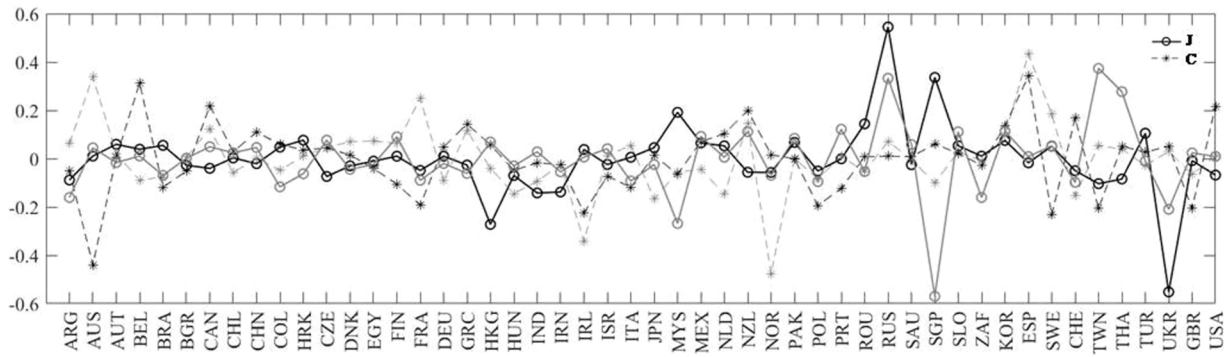
FIG. 8. Components of eigenvectors of **J** (full line) and **C** (dashed line) corresponding to the first two lowest eigenvalues. Eigenvectors associated with the smallest eigenvalues of both **J** and **C** are responsive to the large values of J_{ij} and C_{ij}, respectively. For example the first two largest components of the eigenvector related to the largest eigenvalue of J (black line) correspond to Russia and Ukraine, whose J_{ij} is the largest of the elements of **J**.

product of the eigenvectors of respectively the matrix C and of the matrix J, i.e., $|\xi_i \cdot \zeta_j|$ is taken into account. Figure 9(b) shows the scalar product of couples of eigenvectors related to the two matrices ordered for increasing value of eigenvalues. The block corresponding to largest eigenvalues is nearly diagonal, outlining how for these eigenvalues the eigenvectors decomposition of **C** and **J** is similar. The eigenvectors of **C** and **J** related to lowest eigenvalues, which carry respectively information on the couples of countries strongly interacting and highly correlated, do not preserve such a correlation. This emphasizes that there is not a one-to-one correspondence between largest values of J_{ij} and C_{ij}, see also Fig. 8.

B. Hierarchical clustering and principal components analysis of J

In the following, we apply to the set of inferred interactions $\{J\}$ two general methods usually applied to the analysis of

correlation matrices, i.e., HC and PCA [17,18]. Figures 10(a) and 10(b) show the inferred elements of the matrix **J**. For sake of clarity, the elements of the interaction matrix have been ordered following the HC outputs described in the following.

The HC is a hierarchical clusterization method [45]. First, it is defined a metrics and it is calulated the distance between each two columns of the interaction matrix. In the present case the metrics adopted is city block distance, $d(\bar{J}_i, \bar{J}_j) = \sum_\alpha |J_{i\alpha} - J_{j\alpha}|$, where $\bar{J}_{i(j)}$ states for a column of **J**. At the starting step each column of **J** corresponds to a different cluster. At each step the two clusters at the shortest distance are merged and form a cluster. This protocol allows to build a so-called dendrogram, the tree diagram shown in Figs. 10(b) and 10(d). The height of the link between two objects, i.e., the countries displayed on the horizontal axis, indicates the distance between the objects. The dendogram can be cut at a given height thus allowing the definition of a certain number of clusters. The right panels of Fig. 10 show the result of the
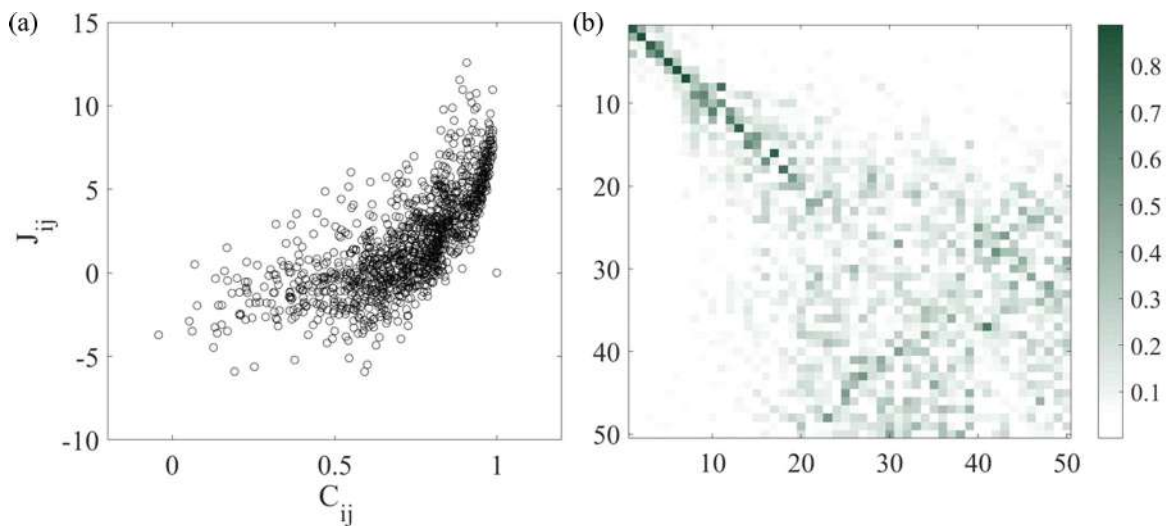

FIG. 9. (a) Interactions J_{ij} plotted against C_{ij}. (b) Matrix reprsentation of the absolute value of the scalar product of the eigenvectors of C and J, $|\xi_i \cdot \zeta_j|$ sorted by descending order of corresponding eigenvalues. While the block corresponding to the largest eigenvalues is almost diagonal, emphasizing that the eigenvectors corresponding to the largest eigenvalues of both **J** and **C** have a similar components decomposition, the block corresponding to the smallest eigenvalues is not. Since the components of eigenvectors corresponding to smallest eigenvalues are related to couples of countries strongly interacting and highly correlated, respectively, this shows that, due to the existence of indirect correlation, there is not a one-to-one correspondence between the largest values of J_{ij} and C_{ij}.
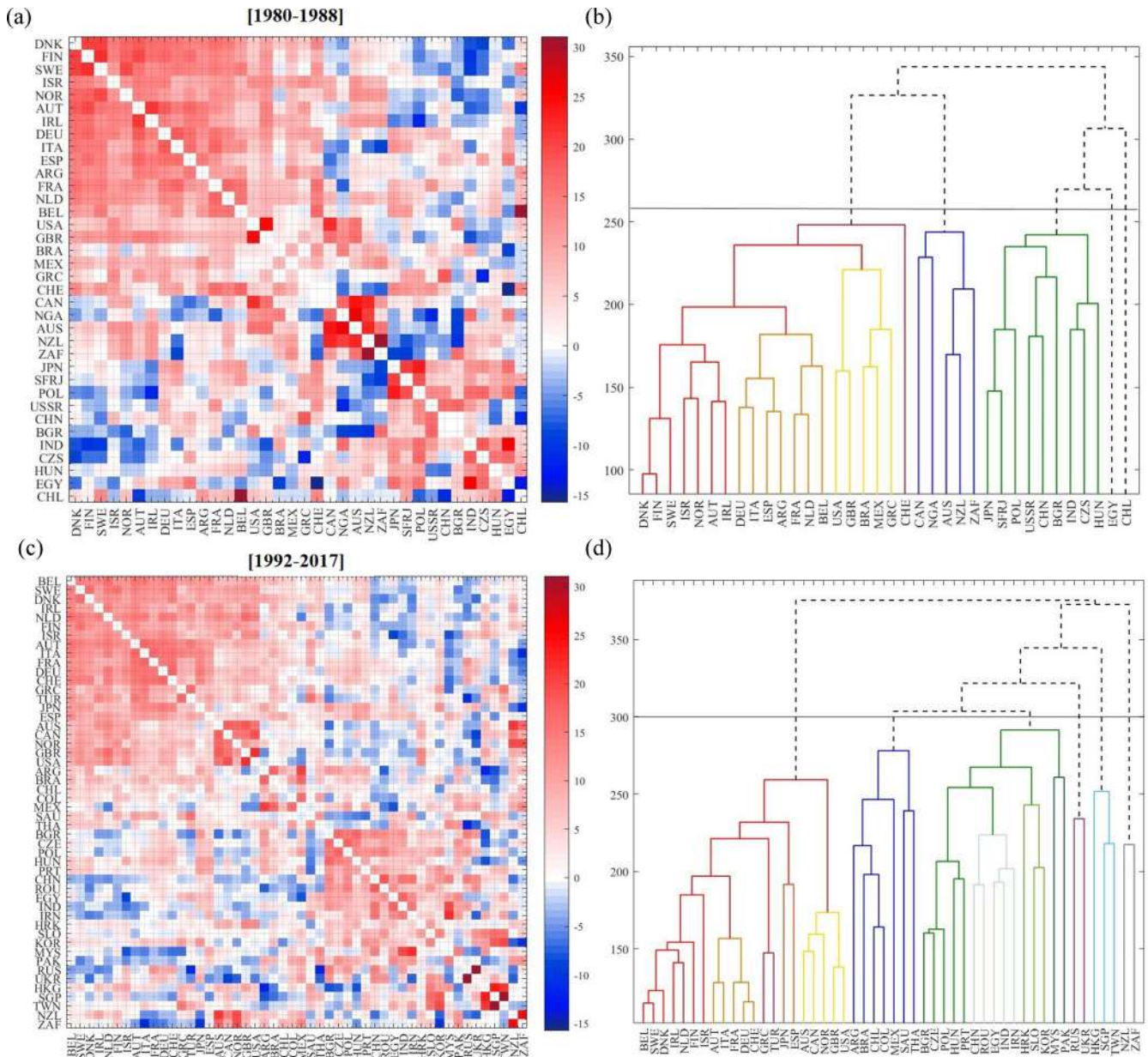
FIG. 10. (a) Inferred interactions J_{ij} in the period 1980–1988. (b) Dendrogram obtained by the hierarchical clustering analysis of **J** in the period 1980–1988. The height of the link between two objects (countries or clusters of countries) is proportional to the distance between the objects. The horizontal line shows where the dendrogram can be cut and clusters identified. (c) Inferred interactions J_{ij} in the period 1992–2017. (d) Hierarchical clustering analysis of **J** in the period 1992–2017. For sake of clarity, the elements of the matrix interactions have been sorted by following the order defined by the hierarchical clustering. The two Western clusters (including the Western Europe countries and the USA) are clearly observable in the matrix interactions. The pairwise interactions between these countries are furthermore all positive, outlining a tendency to alignment.

HC protocol applied to **J**, different colors identify different clusters.

The PCA is a partial eigendecomposition of the matrix **J**, where only the eigenvectors corresponding to the largest eigenvalues are considered [17,18]. The selcted eigenmodes are those whose eigenvalues summed up descirbe with an error small enough the trace of the original matrix. Following this criterion we select the eigenvectors corresponding to larger (in magnitude) eigenvalues of **J**. As discussed in Sec. V such eigenmodes bring genuine information, see also Fig. 6.

A bidimensional representation of the eigenvectors of **J** corresponding to the two largest eigenvalues is obtained by plotting in the bidimensional plane the points whose coordinates are the components of the two corresponding eigenvectors, as shown in Fig. 11. This plot allows the identification of clusters of countries which have a similar interaction with the rest of the whole system [45], as detailed in the following. The value of a given component of the eigenvectors related to the largest eigenvalues bring information on how the country associated to the given component interact with all the other countries.
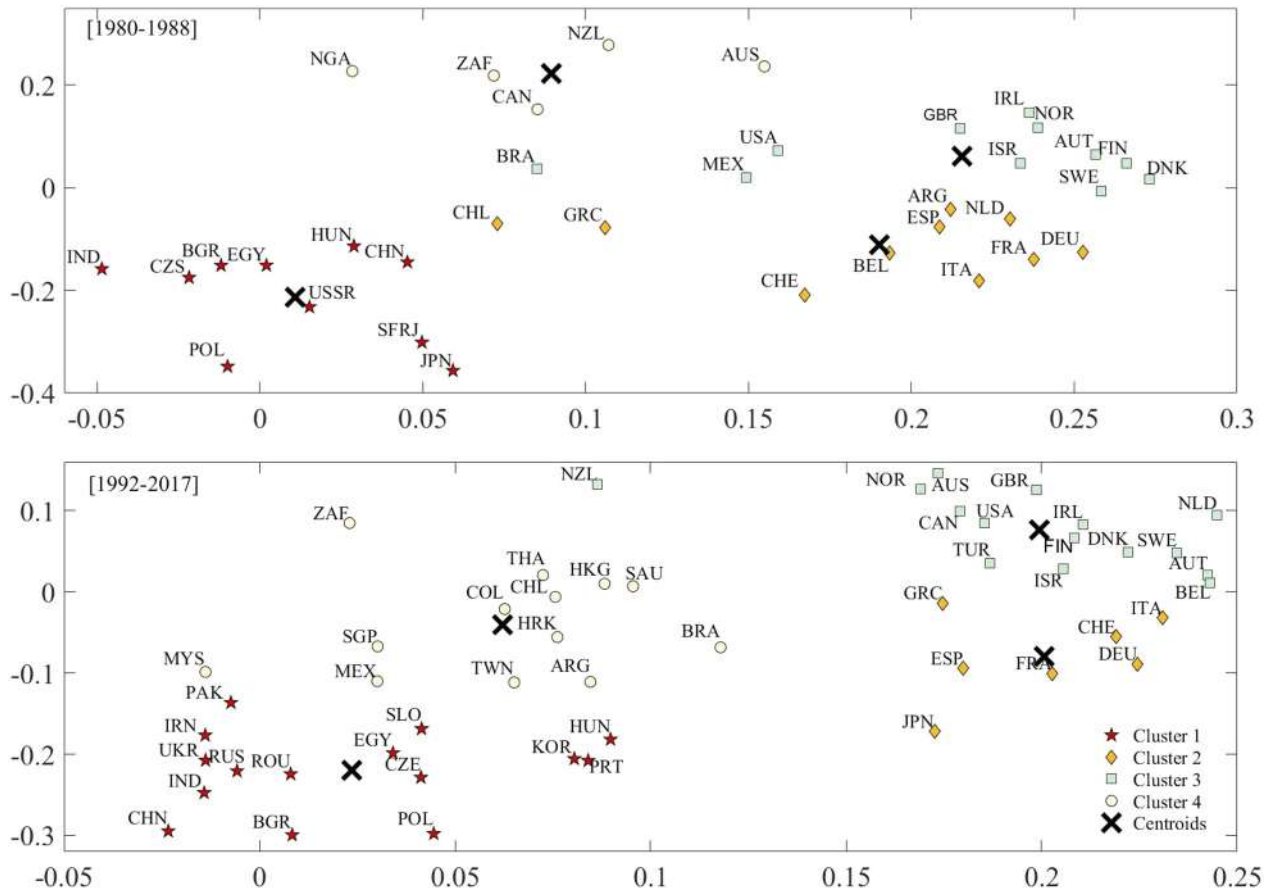
FIG. 11. PCA analysis of {J} in the time intervals 1980–1988 and 1992–2017. 2D graphycal representation of the eigenvectors related to the first two largest eingevalues of **J** is shown. The distance between two countries, following the similarity criterion described in the text, can be thus easily visualized. PCA analysis allows the identification of clusters. Elements belonging to the same cluster are identified by a given symbol and color. Clusters centroids are shown by crosses. A map representation of the PCA clusters is shown in Fig. 12.

For example if the interaction network is such that only the interaction between countries i and j is different from zero, whereas all the others are zero, the eigenvector of the interaction matrix corresponding to the largest eigenvalue will have the only components i and j different from zero. A similarity criterion can be thus established: the two countries associated to the eigenvector's components which have a similar value (they are the only two different from zero) are those which have a similar among them, but different with respect to all the others, network of interactions (they are the only two interacting with at least one other country). For more complex cases, where the interaction matrix has more than one element different from zero, one can assume that those countries which are neighbors on the PCA plane are connected with other nodes with a similar interactions network. Notice that two countries similar following the PCA criterion can have a small pairwise interaction. A K-means clustering procedure with a squared euclidian metrics [17,18] is used in order to obtain the clusters decomposition shown in Fig. 11. The number of clusters has been chosen equal to four. The centroids of the cluster and the attribution of a given point to a cluster are determined by minimizing the function $\chi_K = \sum_{k=1}^{4} \sum_{i \in C_k} d(x_i, c_k)$, where i is the point's index, k indices the cluster C_k and $d(x_i, c_i)$ is the distance between points and centroids in the specified

metrics (euclidean in the present case). $l2$-regularization have been introduced in the optimization routine (parameter 0.13). HC and K-mean clustering have been performed by MATLAB packages [68]. Figure 12 finally shows in a geographical map the clusters obtained by PCA.

VI. A GEOPOLITICAL FEEDBACK

Even if a geopolitical analysis of the results obtained (Figs. 11 and 12) is beyond the aim of the present work, we shortly point out in the following how the inferred results are in agreement with the general lines one can draw basing only on sinple geopolitical arguments concerning international relations in a global context both before and after the fall of the Berlin Wall. In the time interval 1980–1988, the existence of a so-called Communistic block, including Soviet Union, Eastern Europe countries and China (cluster 1 in Fig. 11, top panel) can be clearly observed. Interestingly, India, China, and Japan belong to this same cluster centered in the Soviet Union. The existence of a similar cluster is preserved in the period 1992–2017, following the fall of the Berlin Wall, with the only relevant exception of Japan, which during the more recent time interval belongs to the Western block (cluster 2 in Fig. 11, top panel). Partnership between India,
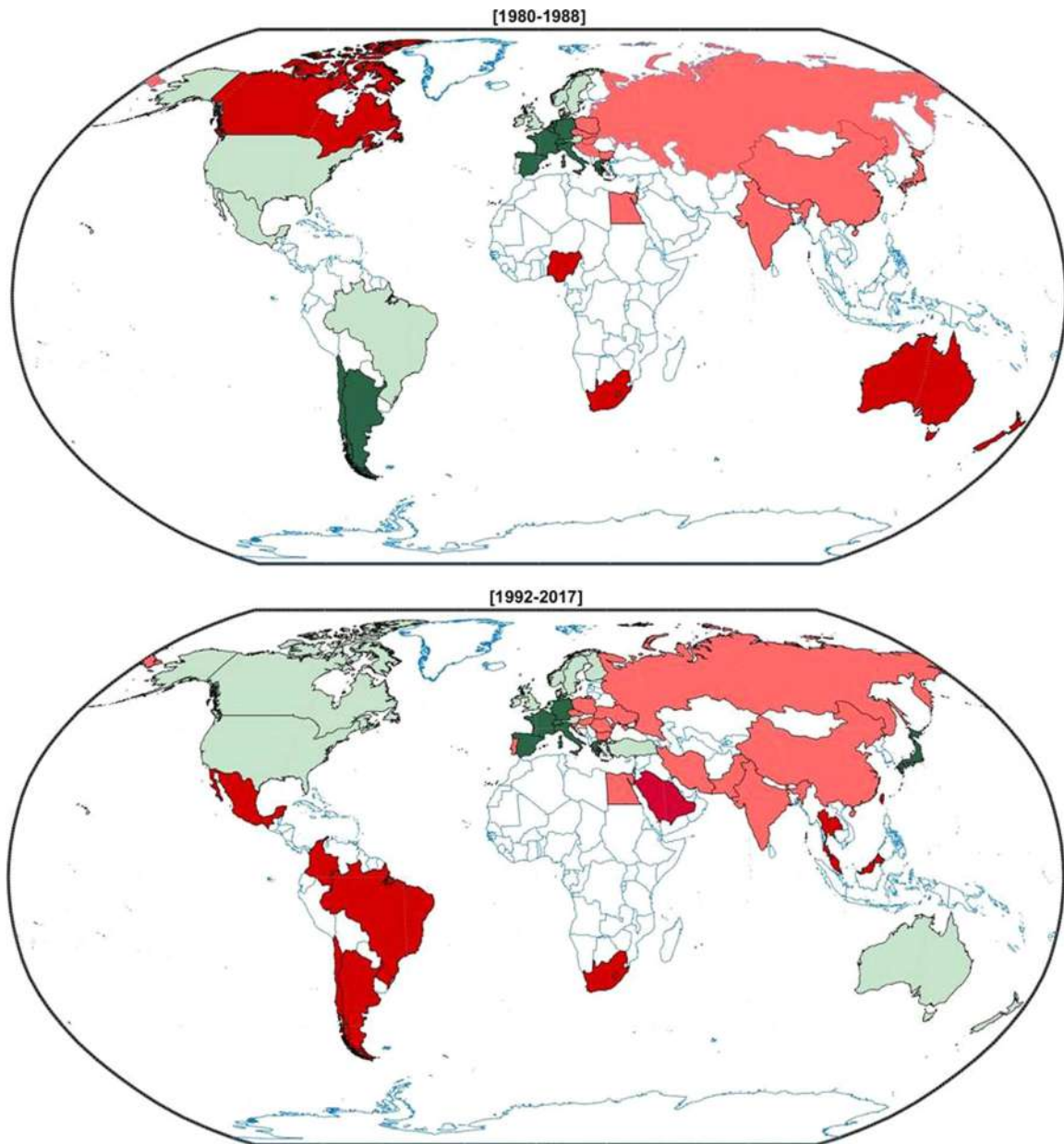
FIG. 12. Map representation of PCA clusters. Different colors identify different clusters.

Soviet Union (or Russia) and China, either during the Cold War and after are well-documented [71–74], also partially formalized in the so-called BRICS alliance [75]. The role of Japan to the Cold War and its position change possibly dating at its involvement to the Gulf war in 1991 [71] and at the establishment of the Japanese-American Security Treaty in 1996 [71,76] is also a well-assessed concept in the historical analysis.

The PCA analysis emphasizes the presence of other two clusters in the time period 1980–1988: one including the Central and South Europe countries (cluster 2) and a second one including United Kindom, United States and Scandina-vian countries (cluster 3 in Fig. 11, top panel). As it is possible to observe in Fig. 11 the centroids of clusters 2 and 3 are closer and they are both distant from the centroid of cluster 1. This bipolar configuration is mostly preserved in the time period

1992–2017, where the presence of one pole encompassing the North American and Europeans countries and the other one the former communist block countries, is still observable, de-spite the fall of the Berlin Wall. This result possibly supports the hypothesis that the conflict, which took shape in the Cold War and was also played on the plane of a territorial control, was anyhow mantained, after the fall of the Berlin Wall, on a sociopolitical, cultural and economic plane. It is also relevant to observe that in the post-Wall period the distance between United State and the Western Europe countries shortened. Furthermore, whereas in the period 1980–1988 the Latin America countries belonged to the same cluster than United States and Western Europe countries, in the more recent time interval they aggregate in a separate cluster (cluster 4 in Fig. 11, bottom panel). This rearrangement is accompanied by the rapprochement of United States to Western Europe

countries, as noticed above, and, similarly, of Canada and Australia, which only in the time-period 1992–2017 belong to the same Western countries cluster (cluster 3 in Fig. 11, bottom panel).

Finally in the following we analyze a bit more in details the European configuration observed in the two time period, in particular putting it in relation with the shaping of the European Community and NATO alliance. The single European Act of 1986 stipulated that by the beginning of 1993 free movement of goods, services, capital and labor among the twelve member states of the so-called European Community (EC) would be achieved. The EC refers to the association of countries from the European Economic Community, the European Coal and Steel Community and the European Atomic Energy Community taken place in 1967 [71]. The original nucleus of EC are the six signatories of the Treaty of Rome (France, Germany, Italy, Belgium, the Netherlands, Luxembourg) plus Great Britain, Ireland, Denmark, Greece, Spain, and Portugal. Accordingly, during the time interval 1980–1988 the six original EC's countries all belong to cluster 2, together with Spain and Greece. Great Britain, Ireland, and Denmark, although next-neighbors of the above-cited EC countries, belong to Cluster 1. Cluster 1 furthermore encompasses the states of the European Free Trade Association (EFTA), i.e., Austria, Finland, Norway, and Sweden. Emerged in 1960 as a rival to the old European Economic Community, EFTA, apart from Norway, merged into the EC in 1995 [71]. Negotiations with the four countries began in 1993. Consistently with this scenario the core of the EC countries and the four EFTA countries belong to different, but next neighbors, clusters (cluster 2 and 3 in Fig. 11, top and bottom panels) in both the periods [1980–1988] and [1992–2017]. The members of NATO after 1992 are the twelve founding members (the United States, the United Kingdom, Belgium, Canada, Denmark, France, Iceland, Italy, Luxembourg, the Netherlands, Germany Norway, Portugal), Greece, Turkey, Spain, the former Warsaw Pact countries (the three Eastern European countries Hungary, the Czech Republic, Poland, ex members of COMECON, accessed in 1997; Bulgaria, Romania, Slovakia, Slovenia and the Baltic states Estonia, Latvia, and Lithuania in 2004), Albania and Croatia entering in 2009 and Montenegro in 2017. As observed above, the inclusion of the Eastern European states in the NATO doesn't correspond to a migration of these states to the Western countries clusters in the time period 1992–2017. This last point gives thought to the relations between NATO and Eastern Europe countries [77].

This brief and essential overview, while not at all exhaustive, aims on one hand to highlight the reasonable agreement observed between the results obtained and an elementary geopolitical analysis and, on the other hand, to outline how this quantitative analysis can be a valuable instrument for a historical analysis.

VII. SUMMARY AND OUTLOOK

We defined the disciplinary profile of a country as a versor whose elements are the number of articles published by the given country in a given discipline divided the total number of articles published by the country. The countries considered are those supporting a significant number of published articles with respect to the worldwide production. Each country can be associated to the node of a graph. A partial definition of cultural production of a given country, i.e., only the one which take shape in articles production and recorded in the specified databases, is adopted. This has to be taken in mind if a historical analysis is done on the basis of the present results. Time series of country-level disciplinary profiles are acquired in the time intervals 1980–1988 and 1992–2017 with a time step of a year. Since in between the two time intervals, the fall of the Berlin Wall, with the related historical events, caused a reorganization of the geopolitical map, the graph under exam would need to be differently defined in the two time intervals. For each graph the set of the pairwise interactions has been inferred. A comparison between the results obtained in the two time intervals, however, can be of interest in a geopolitical perspective. A preliminary analysis of the empirical cross-correlation matrices of the disciplinary profiles in the more recent time interval has been performed. By exploiting a comparison with RMT results it was possible to establish that the cross-correlation matrices contain, beyond noise content, genuine information, stationary in time. After proving that the empirical cross-correlation matrices bring genuine information, an inference procedure based on maximum entropy modeling of second-order marginal has been applied to the data in order to infer the value of pairwise interactions J_{ij}. The maximum entropy modeling is equivalent to the maximization of a likelihood function belonging to the class of Boltzmann distribution related to a generalized Heisenberg model with n_d-dimensional spin variables. In the present case, n_d is equal to the number of disciplines considered. In order to obtain a working algorithm able to draw the optimal matrix of pairwise interactions, we used a pseudo-likelihood maximization approach. We analytically computed the pseudolikelihood and its gradient in order to facilitate the computational solution of the inference problem. The analytical expressions deserve by themselves interest in Bayesian inference framework whenever a n_d-dimensional Heisenberg model is appropriate for the inference problem one aims to solve. We finally obtained the optimal value of the matrix **J** by numerical maximization of the Log-pseudolikelihood function. To the inferred interactions matrix they have been applied two classification methods, Hierarchical Clustering and Principal Component Analysis, usually exploited in the analysis of the cross-correlation matrices. We obtained a clusters representation of the interactions network shown in Figs. 10–12. An elementary geopolitical analysis of the results obtained not only emphasizes the soundness of the results obtained but it calls for deeper historical analysis and, finally, it support the use of physical modeling in this field.

Beyond a geopolitical analysis of worldwide international relationships, which can be carried on by exploiting the presented results and methodology, this study highlights the profound question of what mechanisms drive the collective knowledge and science, in particular, evolution. In Ref. [78], basing on a quantitative analysis of topic words and phrases in titles and abstracts of publications of the American Physical Society, insights on the trends of discoveries in this discipline has been obtained. It emerged that the rise and fall of physical concepts are self-organized and it was possible to identify the

onset of the so-called Matthew effect [79]. As complementary to this study, our work pointed out how, on a lower level of detail because disciplines rather than paradigms in a given discipline are considered, the disciplinary profile of a country can be also driven by exogenous factors such as international geopolitical interactions.

Whereas the present work is aimed to the inference of the structure of the geopolitical network from the observed cross-correlation matrices basing on the maximum entropy principle stated in information theory, which allows to minimize the number of assumptions, some relevant aspects of the phenomenon highlighted in the present study deserve attention and can be the focus of further analyses. Disentangling the role of pairwise interactions between countries from external influences in the shaping of the disciplinary profile is one of these. External influences enclose every possible bias that can influences the disciplinary profile a part from pairwise interactions, e.g., the attachment to a given disciplinary profile model due to cultural and ideological traditions or promoted by different amount of public investments in the academic and research sector. External influences can be modeled by introducing an external field or dummy nodes with directed links on which sit fixed oriented spin variables. Similar approaches have been recently exploited in the extended voter model [80]. A second relevant question concerns the dynamics of the disciplinary profiles. Particularly interesting to this respect would be the analysis of contagion, simulated as a stochastic process, in the multidimensional Heisenberg model. Models able to interpolate between regular and random connection topology, such as the small-world model [81], or to account for the heterogeneous structure of the network on the diffusion dynamics [82], are particularly suitable to deal with this issue.

We finally point out two outlooks of the present work. Correlations among two variables $\mathbf{s}_i$ and $\mathbf{s}_j$ can be caused either by direct statistical coupling and indirect correlation effects, such it is the case, e.g., in the Heisenberg model of two variables $\mathbf{s}_i$ and $\mathbf{s}_j$ not interacting among themselves but both interacting with a third variable $\mathbf{s}_k$. The inference of the interactions matrix $\mathbf{J}$ allows to identify variables statistically coupled. Once the matrix $\mathbf{J}$ has been inferred it could be interesting to disentangle direct and indirect correlations. Given two nodes i and j of a graph this can be achieved by defining the so-called direct information [53] $DI_{ij} = \sum_{\{\mathbf{s}_i, \mathbf{s}_j\}} P_{ij}^{(dir)}(\mathbf{s}_i, \mathbf{s}_j) \ln \frac{P_{ij}^{(dir)}(\mathbf{s}_i, \mathbf{s}_j)}{f_i(\mathbf{s}_i) f_j(\mathbf{s}_j)}$, where the sum is on the two-variables configuration space, $\{(\mathbf{s}_i, \mathbf{s}_j)\}$, and $f_i(\mathbf{s}_i) = \sum_{\mu=1}^{M} \delta(\mathbf{s}_i^\mu - \mathbf{s}_i)$ is the frequency of observation in the time series $\{\mathbf{s}_i^\mu\}$ of the variable value $\mathbf{s}_i$. The so-called two-variable direct probability $P_{ij}^{(dir)}(\mathbf{s}_i, \mathbf{s}_j)$ is the key quantity in DI_{ij}. It can be obtained through the definition of so-called three messages $\nu_{i \to j}(\mathbf{s}_i)$, i.e., $P_{ij}^{(dir)}(\mathbf{s}_i, \mathbf{s}_j) = \frac{\nu_{i \to j}(\mathbf{s}_i) e^{-J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j} \nu_{j \to i}(\mathbf{s}_j)}{\sum_{\{\mathbf{s}_i, \mathbf{s}_j\}} \nu_{i \to j}(\mathbf{s}_i) e^{-J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j} \nu_{j \to i}(\mathbf{s}_j)}$ [53,83]. The message $\nu_{i \to j}(\mathbf{s}_i)$ is the marginal distribution of the variable $\mathbf{s}_i$ in a modified graph which does not include the node i. The message effectively represent the contribution to the correlation between the two variables not attributable to the direct coupling between them. Once $\mathbf{J}$ is inferred, e.g., by exploiting the pseudolikelihood approach as done in the present work,

the three messages can be calculated through a self-consistent procedure, so-called belief propagation algorithm [83].

Inference method based on maximization of the likelihood function in order to infer the interactions J_{ij} can be subject to overfitting due to the small number of collected configurations, i.e., limitness of the time-series. A way to avoid such a trouble is to use low-correlated data whose empirical cross-correlation function has an effective rank high enough [84]. Alternatively, the number of free parameters, i.e., J_{ij} can be reduced, as done, e.g., in the so-called decimation algorithm [85]. In the mean field-approximation an expression of the likelihood as a function of the eigenvalues of the correlation matrix is obtained [86]. It is interesting to consider the explicit result obtained in Ref. [86] for the log-likelihood function, $\ln l = \sum_{i=1}^{N} \sum_{\{\mathbf{s}_i\}} f_i(\mathbf{s}_i) \ln f_i(\mathbf{s}_i) + \frac{1}{2} \sum_n \lambda^{(n)} - 1 - \ln \lambda^{(n)}$, where $\lambda^{(n)}$ is the nth eigenvalue of $\mathbf{C}$. From the expression above it follows that the single-eigenvalue contribution to the log-likelihood can be isolated and that the larger contribution comes from both largest and smallest eigenvalues of $\mathbf{C}$. This is in agreement with the results of Sec. III A showing that these eigenvalues are those bringing the genuine information. It can be thus defined a hierarchy of eigenvalues according to their contribution to the log-likelihood function. To each eigenvalue it is associated an eigenvector, whose nonnull components define in the graph a pattern of nodes. Only pairwise interactions between the nodes belonging to the patterns associated to eigenvalues significantly contributing to the log-likelihood can be then the free parameters. This can give rise to a kind of piloted decimation. The result sketched above, furthermore, emphasizes that in the mean-field approximation to the log-likelihood function they contribute not only the largest eigenvalues of $\mathbf{C}$ but also the smallest. Performing PCA on the cross-correlation matrix allows to identify the path of nodes of maximum covariance. This is, however, not equivalent to find the path of maximally interacting nodes because, it is neglected the contribution of the lowest eigenvalues of $\mathbf{C}$, which instead in the mean field-approximation contributes to the log-likelihood function [86] and could anyhow affect the inferred $\mathbf{J}$. This, together with the fact outlined above that correlations can be either direct or indirect, strengthens the need to go beyond analysis of empirical cross-correlation matrices and instead to infer $\mathbf{J}$ if the aim one points to is to analyze the properties of the underlying interactions network generating the observed correlations.

Authors Contribution: M.G.I did theoretical and numerical computations, conceived, and performed the data analysis and inference. C.D. introduced to the research topic and provided the data. G.R. proposed the modeling. L.L. introduced the pseudolikelihood approach. M.G.I., C.D., L.L., G.Q., and G.R. discussed the results. M.G.I. wrote the paper. C.D. and G.Q. wrote Sec. II. M.G.I., C.D., L.L., G.Q., and G.R. revised the paper.

[1] J. S. Nye Jr., *Soft Power* (PublicAffairs, New York, 2004).

[2] B. Esambert, *La Guerre Économique Mondiale* (Olivier Orban, Paris, 1991).

[3] G. Gagliano, *The Economic Warfare* (Cestudec, Como, Italy, 2018).

[4] http://www.europarl.europa.eu/summits/lis1_en.htm.

[5] Henri Dou, A. Juillet, and P. Clerc, *Strategic Intelligence for the Future 1* (Wiley, London, UK, 2019).

[6] D. Stelowska, Culture in international relations defining cultural diplomacy, Pol. J. Polit. Sci. **1**, 50 (2015).

[7] V. Huseynov, Soft power geopolitics how does the diminishing utility of military power affect the Russia West confrontation over the Common Neighbourhood, East. J. Eur. Stud. **7**, 71 (2016).

[8] R. Cruz and De Castro, Confronting china's charm offensive in east asia: A simple case of fighting fire with fire? Issues & Studies **45**, 71 (2009).

[9] M. O. Jackson, *Social and Economic Networks* (Princeton Univ. Press, Woodstock, 2008).

[10] H. V. Ribeiro, L. G. A. Alves, A. F. Martins, E. K. Lenzi, and M. Perc, The dynamical structure of political corruption networks, J. Complex Netw. **6**, 989 (2018).

[11] M.-G. Hâncean, M. Perc, A. Gheorghita, G. G. Vega Yon, and B.-E. Mihaila, The formation of political discussion networks, R. Soc. Open Sci. **9**, 211609 (2021).

[12] M. W. Feldman and L. L. Cavalli-Sforza, *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton Univ. Press, New York, 1981).

[13] G. Ruocco, C. Daraio, V. Folli, and M. Leonetti, Biliometric indicators: the origin of their log-normal distribution and why they are not a reliable proxy for an individual scholar's talent, Pal. Commun. **3**, 17064 (2017).

[14] M. L. Metha, *Random Matrices* (Academic Press, Boston, 1991).

[15] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, and H. E. Stanley, Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series, Phys. Rev. Lett. **83**, 1471 (1999).

[16] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, T. Guhr, and H. E. Stanley, Random matrix approach to cross-correlations in financial data, Phys. Rev. E **65**, 066126 (2002).

[17] D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, Cambridge, UK, 2012).

[18] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, Unsupervised learning methods for molecular simulation data, Chem. Rev. **121**, 9722 (2021).

[19] Z. Becsi and S. Lahiri, Bilateral war in a multilateral world: carrots and sticks for conflict resolution, Can. J. Econ. **40**, 1168 (2014).

[20] J. E. Garten, *A Cold Peace: America, Japan, Germany, and The Struggle for Supremacy* (Times Books, New York, 1992).

[21] http://help.prod-incites.com/inCites2Live/filterValuesGroup/researchAreaSchema/esiDetail.html.

[22] L. Bornmann and L. Leydesdorf, Macro-indicators of citation impacts of six prolific countries: InCites data and the statistical significance of trends, PLoS ONE **8**, e56768 (2013).

[23] R. M. May, The scientific wealth of nations, Science **275**, 793 (1997).

[24] D. A. King, The scientific impact of nations, Nature (London) **430**, 311 (2004).

[25] L. Y. Yang, T. Yue, J. L. Ding, and T. Han, A comparison of disciplinary structure in science between the G7 and the BRIC countries by bibliometric methods, Scientometrics **93**, 497 (2012).

[26] I. Bongioanni, C. Daraio, and G. Ruocco, A quantitative measure to compare the disciplinary profiles of research systems and their evolution over time, J. Informetr. **8**, 710 (2014).

[27] A. W. Harzing and A. Giroud, The competitive advantage of nations: an application to academia, J. Informetr. **8**, 29 (2014).

[28] S. Radosevic and E. Yoruk, Are there global shifts in the world science base? Analysing the catching up and falling behind of world regions, Scientometrics **101**, 1897 (2014).

[29] P. Albarran, A. Perianes-Rodriguez, and J. Ruiz-Castillo, Differences in citation impact across countries, J. Assoc. Inform. Sci. Technol. **66**, 512 (2015).

[30] P. Lorca and J. de Andrés, The importance of cultural factors in R and D intensity, Cross Cult. Res. **53**, 483 (2019).

[31] T. Pinto and A. A. C. Teixeira, The impact of research output on economic growth by fields of science: A dynamic panel data analysis 1980–2016, Scientometrics **123**, 945 (2020).

[32] C. Daraio, F. Fabbri, G. Gavazzi, M. G. Izzo, L. Leuzzi, G. Quaglia, and G. Ruocco, Assessing the interdependencies between scientific disciplinary profiles at the country level: A pseudolikelihood approach, Scientometrics **116**, 1785 (2018).

[33] C. Daraio, F. Fabbri, G. Gavazzi, M. G. Izzo, L. Leuzzi, G. Quaglia, and G. Ruocco, Assessing the interdependencies between scientific disciplinary profiles at the country level: A pseudo-likelihood approach' in *Proceedings of the 1448 of the ISSI 2017-16th International Conference on Scientometrics and Informetrics* (2017).

[34] M. B. Bostian, C. Daraio, R. Fare, S. Grosskopf, M. G. Izzo, L. Leuzzi, G. Ruocco, and W. L. Weber, Reconstructing nonparametric productivity networks, Entropy **22**, 1401 (2020).

[35] N. D. Tregubova, M. Fabrykant, and A. Marchenko, Countries versus disciplines: comparative analysis of post-soviet transformations in academic publications from Belarus, Russia and Ukraine, Comparat. Sociol. **16**, 147 (2017).

[36] V. A. Markusova, M. Jansz, A. N. Libkind, I. Libkind, and A. Varshavsky, Trends in Russian research output in post-Soviet era, Scientometrics **79**, 249 (2009).

[37] J. Adams and C. King, Global Research Report: Russia. Research and Collaboration in the New Geography of Science, Evidence Ltd, UK (2010).

[38] A. Guskov, D. Kosyakov, and I. Selivanova, Scientometric research in Russia: Impact of science policy changes, Scientometrics **107**, 287 (2016).

[39] S. Shashnov and M. Kotsemir, Research landscape of the BRICS countries: Current trends in research output, thematic structures of publications, and the relative influence of partners, Scientometrics **117**, 1115 (2018).

[40] J. A. S. Almeida, A. Pais, and S. J. Formosinho, Science indicators and science patterns in Europe, J. Informetr. **3**, 134 (2009).

[41] M. Thelwall and J. M. Levitt, National scientific performance evolution patterns: Retrenchment, successful expansion, or overextension, J. Assoc. Inform. Sci. Technol. **69**, 720 (2018).

[42] T. Yue, L. Y. Yang, P. Ahlgren, J. L. Ding, S. Q. Shi, and R. Frietsch, A comparison of citation disciplinary structure in

science between the G7 countries and the BRICS countries, J. Data Inform. Sci. **3**, 14 (2018).

[43] N. Li, Evolutionary patterns of national disciplinary profiles in research: 1996–2015, Scientometrics **111**, 493 (2017).

[44] J. Allik, A. Realo, and K. Lauk, The scientific impact derived from the disciplinary profiles, Front. Res. Metr. **5**, 6 (2020).

[45] P. Barucca, J. Rocchi, E. Marinari, G. Parisi, and F. Ricci-Tersenghi, Cross-correlations of American baby names, Proc. Natl. Acad. Sci. USA **112**, 7943 (2015).

[46] R. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).

[47] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, London Mathematical Society Lecture Note Series 335 (Cambridge University Press, Cambridge, UK, 2006).

[48] R. Speicher, Free convolution and the random sum of matrices, Publ. RIMS. Kyoto Univ. **29**, 731 (1993).

[49] D. Voiculescu, Addition of certain noncommutin random variables, J. Funct. Anal. **66**, 323 (1986).

[50] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, Random-matrix theories in quantum physics: common concepts, Phys. Rep. **299**, 189 (1998).

[51] Y. Roudi, J. Tyrcha, and J. Hertz, Ising model for neural data: Model quality and approximate methods for extracting functional connectivity, Phys. Rev. E **79**, 051915 (2009).

[52] A. Decelle and F. Ricci-Tersenghi, Solving the inverse Ising problem by mean-field methods in a clustered phase space with many states, Phys. Rev. E **94**, 012112 (2016).

[53] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families, Proc. Natl. Acad. Sci. USA **108**, E1293 (2011).

[54] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.023224 for the eigenvectors of **C** and **C**(k) are shown.

[55] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[56] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).

[57] E. Schneidman, S. Still, M. J. Berry, and W. Bialek, Network Information and Connected Correlations, Phys. Rev. Lett. **91**, 238701 (2003).

[58] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, Nature (London) **440**, 1007 (2006).

[59] M. Miotto and L. Monacelli, Entropy evaluation sheds light on ecosystem complexity, Phys. Rev. E **98**, 042402 (2018).

[60] Y. Roudi, E. Aurell, and J. A. Hertz, Statistical physics of pairwise probability models, Front. Comput. Neurosci. **3**, 22 (2009).

[61] J. Besag, On the statistical analysis of dirty pictures, J. R. Stat. Soc. B **48**, 259 (1986).

[62] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE PAMI **PAMI-6**, 721 (1984).

[63] D. M. Greig, B. T. Porteous, and A. H. Seheuly, Exact maximum a posteriori estimation for binary images, J. R. Stat. Soc. B **51**, 271 (1989).

[64] D. Stauffer, Social applications of two-dimensional Ising models, Am. J. Phys. **76**, 470 (2008).

[65] T. Broderick, M. Dudik, G. Tkacik, R. E. Schapire, and W. Bialek, Faster solutions of the inverse pairwise Ising problem, arXiv:0712.2437.

[66] A. Hyvärinen, Consistency of pseudolikelihood estimation of fully visible Boltzmann machines, Neural Comput. **18**, 2283 (2006).

[67] E. Aurell and M. Ekeberg, Inverse Ising Inference Using All the Data, Phys. Rev. Lett. **108**, 090201 (2012).

[68] https://mathworks.com.

[69] P. Ravikumar, M. J. Wainwright, and J. D. Lafferty, High-dimensional Ising model selection using ℓ_1-regularized logistic regression, Ann. Stat. **38**, 1287 (2010).

[70] P. Tyagi, A. Marruzzo, A. Pagnani, F. Antenucci, and L. Leuzzi, Regularization and decimation pseudolikelihood approaches to statistical inference in XY spin models, Phys. Rev. B **94**, 024203 (2016).

[71] W. R. Keylor, *The Twentieth Century World and Beyond*, 5th ed. (Oxford University Press, Oxford, 2010).

[72] J. A. Naik, *Russia's Policy Towards India: from Stalin to Yeltsin* (M D Publications, New Delhi, India, 1995).

[73] V. Mastny, The Soviet Union's Partnership with India, J. Cold War Stud. **12**, 50 (2010).

[74] A. I. Singh, India's relations with Russia and Central Asia, Int. Aff. **71**, 69 (1995).

[75] J. O'Neill, *The Bric Road to Growth* (London Pub Partnership, London, UK, 2013).

[76] S. Feske, The US-japanese security alliance: out of date or highly fashionable? JSTOR **11**, 430 (1997).

[77] L. Paquette, *NATO and Eastern Europe after 2000. Strategic Interactions with Poland, the Czech Republic, Romania and Bulgaria* (Nova Science Publisher, Huntington, New York, 2001).

[78] M. Perc, Self-organization of progress across the century of physics, Sci. Rep. **3**, 1720 (2013).

[79] R. K. Merton, The Matthew effect in science, Science **159**, 56 (1968).

[80] D. Braha and M. A. M. de Aguiar, Voting contagion: Modeling and analysis of a century of U.S. presidential elections, PLoS ONE **12**, e0177970 (2017).

[81] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393**, 440 (1998).

[82] D. Braha, Global civil unrest: Contagion, self-organization, and prediction, PLoS ONE **7**, e48596 (2012).

[83] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).

[84] A. Decelle, F. Ricci-Tersenghi, and P. Zhang, Data quality for the inverse lsing problem, J. Phys. A: Math. Theor. **49**, 384001 (2016).

[85] A. Decelle and F. Ricci-Tersenghi, Pseudolikelihood Decimation Algorithm Improving the Inference of the Interaction Network in a General Class of Ising Models, Phys. Rev. Lett. **112**, 070603 (2014).

[86] S. Cocco, R. Monasson, and M. Weigt, From principal component to direct coupling analysis of coevolution in proteins: low-eigenvalue modes are needed for structure prediction, PLoS Comput. Biol. **9**, e1003176 (2013).