Gradients of O-information: Low-order descriptors of high-order dependencies

T. Scagliarini,[1] D. Nuzzi,[1] Y. Antonacci,[2] L. Faes,[2] F. E. Rosas,[3,4,5,6,7] D. Marinazzo,[8] and S. Stramaglia[1,9]

[1]*Dipartimento Interateneo di Fisica, Universitá degli Studi Aldo Moro, Bari and INFN, 70126 Bari, Italy*
[2]*Dipartimento di Ingegneria, Universitá di Palermo, 90128 Palermo, Italy*
[3]*Department of Informatics, University of Sussex, BN1 9RH Brighton, United Kingdom*
[4]*Centre for Psychedelic Research, Department of Brain Science, Imperial College London, SW7 2AZ London, United Kingdom*
[5]*Centre for Complexity Science, Imperial College London, SW7 2AZ London, United Kingdom*
[6]*Data Science Institute, Imperial College London, SW7 2AZ London, United Kingdom*
[7]*Center for Eudaimonia and Human Flourishing, University of Oxford, OX3 9BX Oxford, United Kingdom*
[8]*Department of Data Analysis, Ghent University, 9000 Ghent, Belgium*
[9]*Center of Innovative Technologies for Signal Detection and Processing (TIRES), Universitá degli Studi Aldo Moro, 70126 Bari, Italy*

(Received 24 June 2022; accepted 22 December 2022; published 19 January 2023)

O-information is an information-theoretic metric that captures the overall balance between redundant and synergistic information shared by groups of three or more variables. To complement the global assessment provided by this metric, here we propose the gradients of the O-information as low-order descriptors that can characterize how high-order effects are localized across a system of interest. We illustrate the capabilities of the proposed framework by revealing the role of specific spins in Ising models with frustration, in Ising models with three-spin interactions, and in a linear vectorial autoregressive process. We also provide an example of practical data analysis on U.S. macroeconomic data. Our theoretical and empirical analyses demonstrate the potential of these gradients to highlight the contribution of variables in forming high-order informational circuits.

DOI: 10.1103/PhysRevResearch.5.013025

I. INTRODUCTION

Network science [1], a field encompassing approaches where complex systems are represented by graphs, has grown tremendously in the last 20 years thanks to the development of powerful computational techniques to tease interdependencies out of data [2,3]. However, despite the great success of this endeavor, some important questions about complex systems cannot be properly addressed by dyadic representations, but require us to take into account higher-order interactions involving more than two elements. Such approaches typically represent systems as hypergraphs that can be studied via topological data analysis [4] to reveal the structure of complex systems of interest.

A complementary line of research focuses on emergent properties related to what the system *does* and characterizes its high-order behavior from observed data identifying an equivalent to hyperedges from the dynamics of the data [5]. A prominent role in this literature is played by the framework of partial information decomposition (PID) [6] and its subsequent developments [7], which exploit information-theoretic tools to evidence high-order dependencies in groups of three or more variables—and the description of their synergistic or redundant nature. In this context, redundancy corresponds to information which can be retrieved from more than one source, while synergy corresponds to statistical relationships that exist in the whole but cannot be seen in the parts [8]: see, e.g., Ref. [9] for an application of these principles in neuroscience, a field where multivariate information theory is seen as a promising tool to advance our understanding of brain and cognition [10]. Another popular computational tool is the O-information Ω [11], which captures the overall balance between redundant and synergistic high-order dependencies in complex systems [12,13], whereas a positive (negative) Ω means that the multiplet of variables at hand is dominated by redundant (synergistic) dependencies. The computational complexity of Ω calculation scales more gracefully with the number of variables than that of PID, making it particularly well suited for practical data analysis.

Crucially, the quest for high-order descriptions of complex systems comes with important computational and conceptual costs, as these representations often grow superexponentially with the system size. Moreover, while coarse-grained measures such as Ω exist, their global nature fails short, not being able to provide a local description of how high-order phenomena are distributed across systems of interest. Hence, there is a urgent need for intermediate approaches that can enable a compact yet meaningful representation of informational multiplets. Indeed, the success of network science rests partially on the availability of metrics (e.g., centrality measures, see Ref. [14]) that quantify the role of specific nodes or links in the system—metrics which are not as immediate to develop and grasp for high-order analyses, in particular, when the high-order links are statistical dependencies.

Here we address this problem by introducing an approach that provides low-order (i.e., univariate and pairwise) descriptors of high-order dependencies in the analyzed system. The proposed approach is based on the gradients of the O-information: Instead of focusing on the O-information *of groups of variables*, we focus on the variation of the O-information *when variables are added to the rest of the system to form these groups*. This provides a more nuanced description of synergistic or redundant informational circuits, in which the role of each variable can be disambiguated. This framework is operationalized by means of the definitions and derivations presented below.

II. GRADIENTS OF O-INFORMATION

The O-information Ω is a signed metric that quantifies the balance between redundant and synergistic interactions within a multivariate system. Specifically, the O-information of a system described by n stochastic variables $\boldsymbol{X}^n = (X_1, \dots, X_n)$ can be calculated as [11]

$$\Omega(\boldsymbol{X}^n) = (n-2)H(\boldsymbol{X}^n) + \sum_{i=1}^{n}\left[H(X_i) - H\left(\boldsymbol{X}^n_{-i}\right)\right], \quad (1)$$

where $\boldsymbol{X}^n_{-i}$ denotes the set of all the variables in $\boldsymbol{X}^n$ but X_i, and H is the Shannon entropy. It has been shown that if $\Omega > 0$ then the statistical dependencies among variables are well explained by collective constraints, which in turn implies that the system is redundancy dominated. In contrast, when $\Omega < 0$ then the dependencies are better explained as patterns that can be observed in the joint state of multiple variables but not in subsets of these; in other words, the system is synergy dominated.

To assess the contribution of a given variable X_i to the informational circuit contained in $\boldsymbol{X}^n$, we propose to calculate its gradient of O-information given by

$$\begin{aligned}\partial_i\Omega(\boldsymbol{X}^n) &= \Omega(\boldsymbol{X}^n) - \Omega\left(\boldsymbol{X}^n_{-i}\right)\\ &= (2-n)I\left(X_i;\boldsymbol{X}^n_{-i}\right) + \sum_{k=1,k\neq i}^{n} I\left(X_k;\boldsymbol{X}^n_{-ik}\right),\end{aligned} \quad (2)$$

where I is the mutual information [15]. The quantity $\partial_i\Omega(\boldsymbol{X}^n)$ captures how much the O-information changes because of adding X_i, hence giving an account of how this variable contributes to the high-order properties of the system. Correspondingly, $\partial_i\Omega(\boldsymbol{X}^n) > 0$ means that X_i introduces mainly redundant information, while $\partial_i\Omega(\boldsymbol{X}^n) < 0$ indicates that it fosters synergistic interdependencies.

A direct calculation shows that the following bounds hold and are tight:

$$-(n-2)\ln|\mathcal{X}| \leqslant \partial_i\Omega(\boldsymbol{X}^n) \leqslant \ln|\mathcal{X}|, \quad (3)$$

where $|\mathcal{X}|$ is the cardinality of the largest alphabet in $\boldsymbol{X}^n$ (a proof of this can be found in Appendix A). The asymmetry between the two bounds has an important consequence: *while redundancy can be only built step by step, synergy can be established more rapidly*. Indeed, adding a variable to a system of size $n-1$ might provide a maximal redundant contribution of $\ln|\mathcal{X}|$, while the maximal synergy that it might lend is $(n-2)\ln|\mathcal{X}|$ — which can be substantial if n is large.

Following a similar rationale to the one that leads to Eq. (2), one can further introduce a second-order descriptor of high-order interdependencies by considering gradients of gradients. In particular, the second-order gradient of a pair of variables X_i and X_j can be defined as

$$\partial_j\partial_i\Omega(\boldsymbol{X}^n) = \partial_i\Omega(\boldsymbol{X}^n) - \partial_i\Omega\left(\boldsymbol{X}^n_{-j}\right). \quad (4)$$

This second-order gradient captures how much the presence of the variable X_j alters the variation of O-information of the system due to the inclusion of X_i. It is direct to verify the symmetry $\partial_i\partial_j\Omega(\boldsymbol{X}^n) = \partial_j\partial_i\Omega(\boldsymbol{X}^n)$; therefore, we simply denote this quantity as $\partial^2_{ij}\Omega(\boldsymbol{X}^n)$.

An interesting property of $\partial^2_{ij}\Omega(\boldsymbol{X}^n)$ is that it can be rewritten as a whole-minus-sum property:

$$\begin{aligned}\partial^2_{ij}\Omega(\boldsymbol{X}^n) &= \left[\Omega(\boldsymbol{X}^n) - \Omega\left(\boldsymbol{X}^n_{-ij}\right)\right]\\ &\quad -\left[\Omega\left(\boldsymbol{X}^n_{-i}\right) - \Omega\left(\boldsymbol{X}^n_{-ij}\right)\right] - \left[\Omega\left(\boldsymbol{X}^n_{-j}\right) - \Omega\left(\boldsymbol{X}^n_{-ij}\right)\right].\end{aligned} \quad (5)$$

In other words, $\partial^2_{ij}\Omega(\boldsymbol{X}^n)$ measures to what degree the variation to the O-information due to the inclusions of both X_i and X_j is more than the sum of the variations one obtains when including them separately. Consequently, $\partial^2_{ij}\Omega(\boldsymbol{X}^n)$ vanishes if the variables X_i and X_j are part of independent informational circuits.

The second-order gradient $\partial^2_{ij}\Omega(\boldsymbol{X}^n)$ can be compared with the *local O-information* between the variables X_i and X_j (introduced in Ref. [11]) $I(X_i;X_j;\boldsymbol{X}^n_{-ij})$, which corresponds to the interaction information [16] between X_i, X_j and the variables in $\boldsymbol{X}^n_{-ij}$. Interestingly, for $n = 3$ the local O-information and $\partial^2_{ij}\Omega(\boldsymbol{X}^n)$ coincide, while for $n \geqslant 4$ they generally differ. A key difference between these quantities is that the local O-information treats the rest of the system (i.e., $\boldsymbol{X}^n_{-ij}$) as a whole, while in the value of the former is actually dependent on the specific partition that divides $\boldsymbol{X}^n_{-ij}$ into parts, which gives it more sensitivity to evaluate informational circuits.

Successive gradients follow automatically, resulting in a simple chain rule. If γ is a subset of $\{1, \dots, n\}$ of cardinality $|\gamma|$, then

$$\partial^{|\gamma|}_{\gamma}\Omega(\boldsymbol{X}^n) = \sum_{\alpha\subseteq\gamma}(-1)^{|\alpha|}\Omega\left(\boldsymbol{X}^n_{-\alpha}\right), \quad (6)$$

the sum being over all the subsets α of γ. For example, for triplets of variables, the gradient of the O-information reads

$$\begin{aligned}\partial^3_{ijk}\Omega(\boldsymbol{X}^n) &= \Omega(\boldsymbol{X}^n) - \Omega\left(\boldsymbol{X}^n_{-i}\right) - \Omega\left(\boldsymbol{X}^n_{-j}\right) - \Omega\left(\boldsymbol{X}^n_{-k}\right)\\ &\quad + \Omega\left(\boldsymbol{X}^n_{-ij}\right) + \Omega\left(\boldsymbol{X}^n_{-ik}\right) + \Omega\left(\boldsymbol{X}^n_{-jk}\right) - \Omega\left(\boldsymbol{X}^n_{-ijk}\right),\end{aligned} \quad (7)$$

and measures the irreducible contribution to the O-information by the triplet $\{i, j, k\}$, which cannot be ascribed to the inclusion of pairs nor single variables of the triplet. The potential of interpreting these quantities in a topological manner, as has been done with the entropy [17], is an interesting avenue for future research. Here we are addressing the problem of a low-order representation of complex systems and consequently we limit to consider just the first terms of the expansion above.

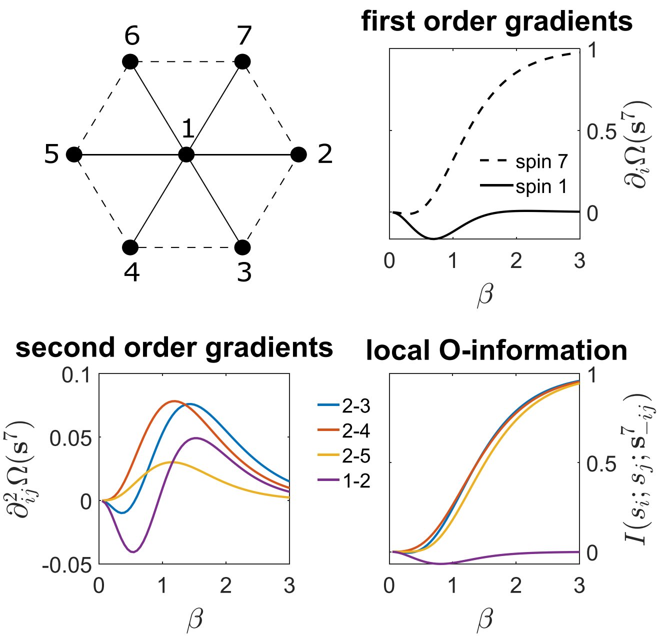

FIG. 1. Top left: The hexagonal geometry of the Ising model, where continuous and dashed lines indicate ferromagnetic and antiferromagnetic interactions, respectively. Top right: The gradients with respect to single spins $\partial_i\Omega(\boldsymbol{s}^7)$ are plotted versus the inverse temperature β. Due to symmetry, the curves of peripheral spins 2–7 are equal. Bottom left: The second-order gradients $\partial^2_{ij}\Omega(\boldsymbol{s}^7)$ are plotted versus β. Due to symmetry, only four nonequivalent curves are plotted. Bottom right: The local O-information $I(s_i;s_j;\boldsymbol{s}^7_{-ij})$ is plotted for the same pairs of spins.

III. PROOF OF CONCEPT

A. Frustrated Ising model

To illustrate the power of the proposed tools, let's start considering an Ising model with Hamiltonian given by

$$\mathcal{H}(\boldsymbol{s}^n) = -\sum_{i\neq j} J_{ij}s_is_j. \tag{8}$$

Our analysis considers a case where $n = 7$, with couplings $J_{ij} = \pm 1$ as depicted in Fig. 1 (top-left panel). Informational measures can be computed directly from the probability of a spin configuration $\boldsymbol{s}^7 = (s_1, s_2, \dots, s_7)$, which is given by

$$p(\boldsymbol{s}^7) = \frac{e^{-\beta\mathcal{H}(\boldsymbol{s}^7)}}{Z}, \tag{9}$$

where Z is the partition function,

$$Z = \sum_{s_1,\cdots,s_7} e^{-\beta\mathcal{H}(\boldsymbol{s}^7)}, \tag{10}$$

and β is the inverse temperature. An evaluation of $\partial i\Omega(\boldsymbol{s}^7)$ shows that the contribution of peripheral spins is dominated by redundancy, while the central spin (the one that, when added to the rest of the system, introduces frustration) introduces synergistic dependencies—with the synergy peaking at a specific temperature. These findings confirm the relationship between synergy and frustration in spin systems already noticed in Ref. [18] while explaining which elements are most responsible for it. Additionally, an analysis using second-order gradients shows that all the pairwise descriptors are associated with redundancy when the temperature is low, while for high temperatures they are synergistic for pairs consisting of the central spin and a peripheral spin, as well as for the pairs consisting of two neighboring spins on the periphery. Overall, these findings add an important spatial description to the previously reported relationship between synergy and higher temperature systems [19]. We note that these findings cannot be retrieved by applications of the local O-information on the same system (Fig. 1, bottom-right panel).

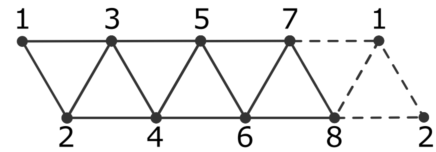

FIG. 2. The triangular geometry of the Ising model with three-spin interactions with periodic conditions in one dimension (indicated with dashed lines).

B. Ising model with three-spin interactions

Here we study a ferromagnetic Ising model of eight spins with the triangular geometry shown in Fig. 2 and three spin interactions, with Hamiltonian

$$\mathcal{H}_3(\boldsymbol{s}^8) = -\sum_{\Delta_{ijk}} s_is_js_k, \tag{11}$$

where the sum runs over all the equilateral triangles. In the thermodynamic limit, this model undergoes a phase transition and it is exactly solvable [20]. In the finite-size system at hand, we may associate a *critical* state to the peak of the magnetic susceptibility, at a finite temperature, see Fig. 3 (bottom right). Considering the gradient with respect to single spins (top left), it shows synergy in anticipation of the critical state, becoming redundant at criticality: similar behavior has been observed for the ferromagnetic pairwise Ising on the square lattice [19]. A similar trend is observed for the links connecting neighboring spins on the lattice (top right), as well as for the triplets of spins belonging to a triangle of the lattice (bottom left): it seems that this peculiar pattern, consisting of synergy anticipating redundancy as the temperature is lowered, is observed if the involved spins form an interaction term of the model's Hamiltonian. For other pairs of spins, and triplets of spins, the pattern of gradients versus temperature is different.

C. Multivariate linear system

As a further simulated system, we consider ten Gaussian processes whose dynamics and interactions are defined by the ten-variate VAR model:

$$\begin{aligned}
Y_{1,n} &= 2\rho_1\cos(2\pi f_1)Y_{1,n-1} - \rho_1^2 Y_{1,n-2} + \epsilon_{1,n},\\
Y_{2,n} &= 0.5Y_{1,n-1} + \epsilon_{2,n},\\
Y_{3,n} &= 0.5Y_{2,n-1} + \epsilon_{3,n},\\
Y_{4,n} &= -0.5Y_{1,n-2} + 0.2Y_{3,n-1} + 0.5Y_{10,n-1} + \epsilon_{4,n},
\end{aligned}$$

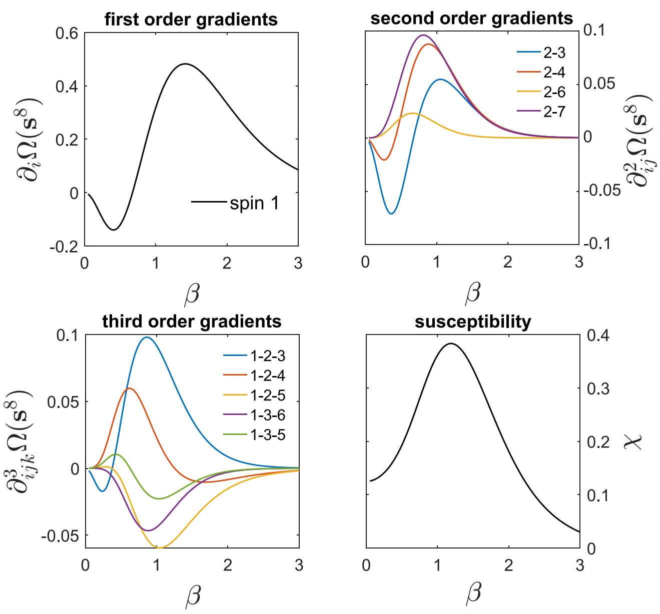

FIG. 3. Top row: The triangular geometry of the Ising model with three-spin interactions with periodic conditions (left) and the magnetic susceptibility of the system (right) as a function of β. Middle row: The gradients with respect to single spins $\partial_i\Omega(\boldsymbol{s}^8)$ (left) and second-order gradients $\partial^2_{ij}\Omega(\boldsymbol{s}^8)$ are plotted versus β (right). Bottom row: The third-order gradients $\partial^3_{ij}\Omega(\boldsymbol{s}^8)$ are plotted versus β (left) and the local O-information $I(s_i; s_j; \boldsymbol{s}^8_{-ij})$ is plotted for the same pairs of spins (right). Due to symmetry, only the nonequivalent curves are shown.

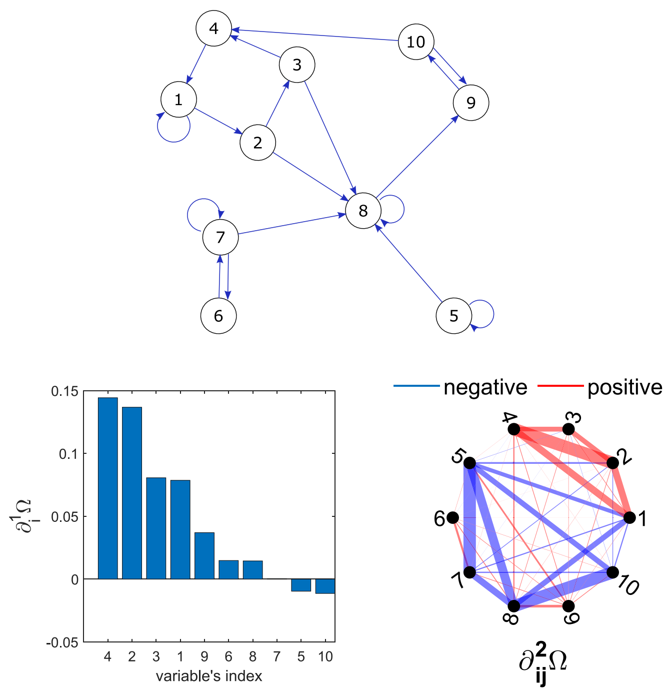

FIG. 4. Top: Visual schematic of the VAR model described by Eqs. (12). Bottom left: First-order gradients $\partial_i\Omega(\boldsymbol{s}^{10})$ of the ten variables. Bottom right: Second-order gradients $\partial^2_{ij}\Omega(\boldsymbol{s}^{10})$ for each pair of variables. The positive and negative networks are normalized separately.

$$
\begin{aligned}
Y_{5,n} &= 2\rho_1\cos(2\pi f_1)Y_{5,n-1} - \rho_1^2 Y_{5,n-2} + \epsilon_{5,n},\\
Y_{6,n} &= 0.3Y_{7,n-2} + \epsilon_{6,n},\\
Y_{7,n} &= 2\rho_1\cos(2\pi f_1)Y_{7,n-1} - \rho_1^2 Y_{7,n-2},\\
&\quad + 0.3Y_{6,n-1} + \epsilon_{7,n},\\
Y_{8,n} &= 2\rho_2\cos(2\pi f_2)Y_{8,n-1} - \rho_2^2 Y_{8,n-2} + 0.4Y_{2,n-2}\\
&\quad + 0.3Y_{3,n-1} - 0.4Y_{5,n-1} + 0.3Y_{7,n-1} + \epsilon_{8,n},\\
Y_{9,n} &= 0.7Y_{8,n-1} - 0.2Y_{10,n-2} + \epsilon_{9,n},\\
Y_{10,n} &= 0.4Y_{9,n-1} + \epsilon_{10,n},
\end{aligned}
\tag{12}
$$

where $\{\epsilon\}$ are independent and identically distributed (i.i.d.) Gaussian noise terms with unit variances $\rho_1 = 0.9$, $\rho_2 = 0.8$, $f_1 = 0.1$, and $f_2 = 0.25$. The network of interactions in this system is depicted in Fig. 4, top; note that variables Y_1, Y_5, and Y_7 represent oscillators with a period equal to $f_1^{-1} = 10$ time steps, while Y_8 is a faster oscillator with period $f_2^{-1} = 4$. This model has been proposed in Ref. [21] to mimic neurophysiological signals, and the high order dependencies of circuits of three and more variables have been analyzed in terms of the O-information rate. However, the role played by single variables or pairs of variables in informational circuits could not be assessed in Ref. [21]—and this is exactly what we do here using gradients of the O-information.

Since we wish to compare the outcomes from our analysis with the known structure of the system, it is worth recalling the notions of mediators, counfounders, and colliders, which arise in the framework of third variable effects [22]. The case where a third variable acts as a mediator (X → mediator ← Y), as well as when the third variable acts as a confounder (X ← confounder → Y), can be associated with redundancy. The case of a collider (X → collider ← Y) is, instead, associated with synergistic effects. In Fig. 4, bottom left, the first-order gradients show that variables Y_1, Y_2, Y_3, and Y_4 are the most redundant with the rest of the system: this is a consequence of the loop of length four which passes through these variables in the graph of interactions, which amounts to mediations among variables. On the other hand, the variables Y_5 and Y_{10} are synergistic: both variables belong to collider subgraphs of interactions, and colliders are known to lead to synergy. Concerning 2-gradients, we find both redundant and synergistic pairs of variables. Comparing with the graph representing the interactions of the system, all the synergistic gradients can be related to colliders of the graph, while redundant gradients can be related to mediators or confounders. What is remarkable is that there are nodes (or links) which are related to colliders and do not show an intense synergy, conversely there are nodes (or links) which belong to chains (or counfounders cliques) which are only slightly redundant. In other words, it is very difficult to foresee which are the synergistic variables of the system just by inspection of the graph of interactions; indeed, the presence of loops implies a cooperative behavior that renders some nodes (or some links) more important for the dynamics of information during the evolution of the system. It is worth mentioning that the frequencies of oscillators also play a role in shaping the information flow pattern in this system.

TABLE I. Gradients of O-information for U.S. macroeconomic indicators (only statistically significant values).

U.S. macroeconomics indicators	$\partial_i\Omega$
COE	0.59
HOANBS	0.47
GDPDEF	0.33
UNRATE	0.27
FEDFUNDS	0.15
TB3MS	0.11
M2SL	0.09
GPDI	−0.26

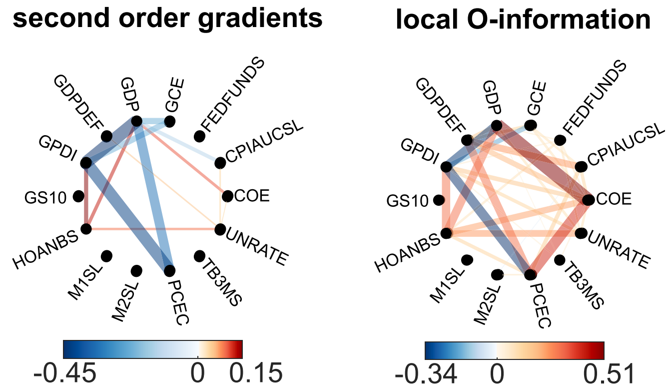

FIG. 5. Left: Second-order gradients for pairs of economic indicators. Right: local O-information of pairs of economic indicators. Edge values are encoded by color (sign) and width (absolute value). Only statistically significant edges—calculated via bootstrap resampling—are included.

IV. APPLICATION TO U.S. ECONOMY

As an econometric application, let's consider 14 U.S. macroeconomic time series taken from the Federal Reserve Economic Dataset (FRED) [23]. We consider quarterly indicators over a period of 61 years (April 1959–January 2020) for a total of 244 observations: paid compensation of employees (COE), consumer price index (CPIAUCSL), effective federal funds rate (FEDFUNDS), government consumption expenditures and investment (GCE), gross domestic product (GDP), gross domestic product price deflator (GDPDEF), gross private domestic investment (GPDI), ten-year treasury bond yield (GS10), nonfarm business sector index of hours worked (HOANBS), M1 money supply (narrow money M1SL), M2 money supply (broad money M2SL), personal consumption expenditures (PCEC), three-month treasury bill yield (TB3MS), and unemployment rate (UNRATE). A wide literature (see, e.g., Refs. [24,25]) has tried leveraging similar data to address the fundamental question regarding the source of economic fluctuations. Here we are neither interested in the role played by shocks and frictions nor in predicting business cycles; rather, our goal is to evidence high-order dependencies in macroindicators of the U.S. economy. To deal with stationary time series, the proposed approach has been applied to the logarithmic returns of the series, over which the gradients of the O-information were calculated using the Gaussian Copula approach described in Ref. [26]. For each gradient, significance testing is performed via bootstrap sampling with replacement: If the 95% confidence interval of that gradient (here computed on 1000 realizations) does not contain zero, the gradient is declared significant. Our results show that seven indicators are redundant with the rest of the system (see Table I—which is consistent with the prevalence of redundancy in real-world multivariate systems, as reflected by latent factors being typically associated to positive O-information (see also Ref. [27]). In contrast, GPDI was found to play a major synergistic role with respect to the rest of the system, which may be associated with the fact that GPDI is considered a good predictor of the productive capacity of the economy.

These first-order analysis can be enriched by the second-order gradient. Results show that several pairs of variables are involved in informational circuits, as shown in Fig. 5: While first-order analysis shows a prevalence of redundancy, second-order gradients show a prevalence of synergy. The most connected node for synergy is GDP, displaying significantly negative $\partial^2_{ij}\Omega$ with four other nodes. When compared with the local O-information (Fig. 5, right), the proposed pairwise descriptors lead to a more sparse and parsimonious pattern.

V. CONCLUSIONS

In this paper, we have introduced the gradients of O-information as measures of how much, in a network of interacting variables, a variable—or a pair of variables—are functionally connected with the rest of a network through redundant and synergistic informational circuits. The use of this metric, together with measures of pairwise dependencies and global measures of higher-order information like regular O-information, provides a more complete description of the informational character of dynamics in complex systems. To illustrate these ideas, we analyzed two small-size Ising systems which allow exact calculation of the informational quantities, thanks to their small state spaces. The first model refers to pairwise interactions and shows that gradients provide a spatial description of high-order dependencies, while confirming the relation between synergy and frustration in spin systems. The second model displays three spins interactions without frustration, and shows that the gradient of single spins shows synergy anticipating the transition of the system toward the ordered state.

Furthermore, we also analyzed a linear multivariate process with assigned directed interactions. Our results show that it is nontrivial to foresee the high-order dependencies from the graph of (pairwise) interactions. In particular, while *a posteriori*—limiting to third variable effects—one may ascribe the informational character of nodes and links to the presence of colliders, confounders, and mediators in the network; in practice, the resulting flow of information is a collective phenomenon which happens to be influenced by the whole network of interactions.

By analyzing the dynamical interactions between U.S. macroeconomics indicators, we have shown how these tools are capable of revealing high-order informational circuits, which evidenced the synergistic role of GPDI—it is a matter for further research to verify if existing econometric models are able to reproduce this high-order behavior exhibited by data.

It is worth stressing that it is not possible to compare the strength of high-order terms (at the same order or across orders); in other words, it is not possible, in general, to state that a given contribution is *more important* than another: All the terms, with their magnitude, contribute to characterize the observed state of the system and gradient terms could be seen as a fingerprint of the observed dynamics of the system. Analogously, in principle, there is not a hierarchy between gradients: each order of gradient measures irreducible effects (i.e., not arising from lowest orders) therefore *a priori* we cannot expect, e.g., that pairwise contributions are more intense than three-variable ones. Depending on the application, synergistic effects may show up at the level of first-order gradients or second-order gradients, and so on. The extension of these measures of high-order behavior to dynamical systems scenarios is also in order, as that would open avenues for investigating the function of complex networked systems on a wide range of applications.

ACKNOWLEDGMENT

T.S., L.F., and S.S. were supported by MIUR Project No. PRIN 2017WZFTZP: Stochastic Forecasting in Complex Systems.

APPENDIX A: BOUNDS FOR FIRST ORDER GRADIENTS OF THE O-INFORMATION

Here we present the proof of the bounds in Eq. (3) in the main text. Let us consider n random variables $\boldsymbol{X}^n = (X_1, X_2, \dots, X_n)$. Two popular extensions of the mutual information are the total correlation $\mathcal{T}(\boldsymbol{X}^n)$ and the dual total correlation $\mathcal{D}(\boldsymbol{X}^n)$,

$$\mathcal{T}(\boldsymbol{X}^n) = \sum_{i=1}^{n} H(X_i) - H(\boldsymbol{X}^n), \tag{A1}$$

$$\mathcal{D}(\boldsymbol{X}^n) = H(\boldsymbol{X}^n) - \sum_{i=1}^{n} H\big(X_i \big| \boldsymbol{X}^n_{-i}\big), \tag{A2}$$

where $H(\cdot)$ is the Shannon entropy. The O-information Ω of the system is given by the difference:

$$\begin{aligned}\Omega(\boldsymbol{X}^n) &\equiv \mathcal{T}(\boldsymbol{X}^n) - \mathcal{D}(\boldsymbol{X}^n) &\text{(A3)}\\ &= (n-2)H(\boldsymbol{X}^n) + \sum_{k=1}^{n}\big[H(X_j) - H\big(\boldsymbol{X}^n_{-j}\big)\big]. &\text{(A4)}\end{aligned}$$

The gradient of the O-information (see the main text) is given by

$$\begin{aligned}\partial_i\Omega(\boldsymbol{X}^n) &= \Omega(\boldsymbol{X}^n) - \Omega\big(\boldsymbol{X}^n_{-i}\big) &\text{(A5)}\\ &= (2-n)I\big(X_i;\boldsymbol{X}^n_{-i}\big) + \sum_{k=1}^{n-1} I\big(X_k;\boldsymbol{X}^n_{-ik}\big). &\text{(A6)}\end{aligned}$$

Analogously, we can consider the gradient of the total correlation

$$\partial_i\mathcal{T} = \mathcal{T}(\boldsymbol{X}^n) - \mathcal{T}\big(\boldsymbol{X}^n_{-i}\big) = I\big(X_i;\boldsymbol{X}^n_{-i}\big) \geqslant 0, \tag{A7}$$

which, being equivalent to a mutual information, satisfies $0 \leqslant \partial_i\mathcal{T} \leqslant \ln|\mathcal{X}|$, where $|\mathcal{X}|$ is the cardinality of the largest alphabet in $\boldsymbol{X}^n$. On the other hand, the gradient of the dual total correlation can be written as a sum of conditional mutual information terms:

$$\begin{aligned}\partial_i\mathcal{D}(\boldsymbol{X}^n) &= \mathcal{D}(\boldsymbol{X}^n) - \mathcal{D}\big(\boldsymbol{X}^n_{-i}\big)\\ &= \sum_{k=1}^{n-1}\big[H\big(\boldsymbol{X}^n_{-i}\big) - H\big(\boldsymbol{X}^n_{-ki}\big) - H(\boldsymbol{X}^n) + H\big(\boldsymbol{X}^n_{-k}\big)\big]\\ &= \sum_{k=1}^{n-1}\big[I\big(X_i;\boldsymbol{X}^n_{-i}\big) - I\big(X_i;\boldsymbol{X}^n_{-ki}\big)\big]\\ &= \sum_{k=1}^{n-1}\big[H\big(X_k\big|\boldsymbol{X}^n_{-jk}\big) - H\big(X_k\big|\boldsymbol{X}^n_{-k}\big)\big] &\text{(A8)}\\ &= \sum_{k=1}^{n-1} I\big(X_k;X_j\big|\boldsymbol{X}^n_{-jk}\big) \geqslant 0, &\text{(A9)}\end{aligned}$$

thus implying that $0 \leqslant \partial_i\mathcal{D}(\boldsymbol{X}^n) \leqslant (n-1)\ln|\mathcal{X}|$.

Putting these results together, one can find that

$$-(n-2)\ln|\mathcal{X}| \leqslant \partial_i\Omega(\boldsymbol{X}^n) \leqslant \ln|\mathcal{X}|. \tag{A10}$$

The reminding of the proof demonstrates these bounds and their tightness.

The upper bound is trivial; indeed, we have already shown that $\partial_i\mathcal{D}(\boldsymbol{X}^n) \geqslant 0$, hence

$$\partial_i\Omega(\boldsymbol{X}^n) = \partial_i\mathcal{T}(\boldsymbol{X}^n) - \partial_i\mathcal{D}(\boldsymbol{X}^n) \leqslant \partial_i\mathcal{T}(\boldsymbol{X}) \leqslant \ln|\mathcal{X}|. \tag{A11}$$

The tightness of the upper bound can be proven by showing that is achieved by the n-`COPY` gate, specifically by taking X_1 as a Bernoulli variable with $p = 1/2$ and $X_1 = X_2 = \dots = X_n$. Since we have that $I(X_i;\boldsymbol{X}^n_{-i}) = 1$ and $I(X_i;\boldsymbol{X}^n_{-ik}) = 1$ for $i = 1, 2, \dots, n$, using Eq. (A5) it follows that

$$\partial_i\Omega(\boldsymbol{X}) = (2-n) + (n-1) = 1, \quad i = 1, 2, \dots, n. \tag{A12}$$

This covers the case of binary random variables, but the result can be readily generalized to $|\mathcal{X}| > 2$.

To prove the lower bound is a little more tricky: We start noting that $\partial_i\Omega(\boldsymbol{X}^n)$ can be written as a sum of conditional interaction information terms. Indeed it has been shown in Ref. [11] that O-information can be decomposed as a sum of interaction information terms

$$\Omega(\boldsymbol{X}^n) = \sum_{k=2}^{n-1} I\big(X_k;\boldsymbol{Y}_1^{k-1};\boldsymbol{Y}^n_{k+1}\big), \tag{A13}$$

where we used the notation $\boldsymbol{Y}^q_k = (X_k, \dots, X_q)$. For definiteness, we fix $i = 1$, obtaining

$$\begin{aligned}\partial_1\Omega(\boldsymbol{X}^n) &= \Omega(\boldsymbol{X}^n) - \Omega\big(\boldsymbol{X}^n_{-1}\big)\\ &= \sum_{k=2}^{n-1} I\big(X_k;\boldsymbol{Y}_1^{k-1};\boldsymbol{Y}^n_{k+1}\big) - \sum_{k=3}^{n-1} I\big(X_k;\boldsymbol{Y}_2^{k-1};\boldsymbol{Y}^n_{k+1}\big)\\ &= I\big(X_2;X_1;\boldsymbol{Y}^n_3\big) + \sum_{k=3}^{n-1}\big[I\big(X_k;\boldsymbol{Y}_1^{k-1};\boldsymbol{Y}^n_{k+1}\big)\\ &\quad - I\big(X_k;\boldsymbol{Y}_2^{k-1};\boldsymbol{Y}^n_{k+1}\big)\big]\\ &= \sum_{k=2}^{n-1} I\big(X_k;X_1;\boldsymbol{Y}^n_{k+1}\big|\boldsymbol{Y}_2^{k-1}\big). &\text{(A14)}\end{aligned}$$

Now we notice that each term in the sum can be written as a difference of two conditional mutual information terms (bounded between 0 and $\ln|\mathcal{X}|$), hence each term has the following bounds:

$$-\ln|\mathcal{X}| \leqslant I\left(X_k; X_1; \boldsymbol{Y}_{k+1}^n \middle| \boldsymbol{Y}_2^{k-1}\right) \leqslant \ln|\mathcal{X}|. \tag{A15}$$

This implies that

$$-(n-2)\ln_2|\mathcal{X}| \leqslant \partial_1\Omega(\boldsymbol{X}^n), \tag{A16}$$

thus proving the lower bound. Finally, to prove that the lower bound is tight, we consider the n-XOR gate, that is, $X_1 \dots X_{n-1}$ as Bernoulli random variables with $p = 1/2$ and $X_n = (\sum_{j=1}^{n-1} X_j) \bmod 2$. Using Eq. (A5), we have $I(X_i; \boldsymbol{X}_{-i}^n) = 1$ and $I(X_k; \boldsymbol{X}_{-ik}^n) = 0$; then

$$\partial_i\Omega = (2-n), \quad i = 1, 2, \dots, n. \tag{A17}$$

Then, the lower bound is tight, since is achieved by the variables composing a n-XOR gate.

APPENDIX B: O-INFORMATION OF U.S. MACROECONOMIC INDICATORS: TRIPLETS AND QUADRUPLETS

Concerning the U.S. economic data set, we report here the conventional O-information analysis, taking into account triplets and quadruplets of variables. We first obtain all the triplets which are significantly synergistic and those which are significantly redundant. Then, for each variable, we sum Ω over all the redundant and significant triplets which contain that variable, obtaining R_Ω which is an index of redundancy of that variable. The same is done summing over synergistic triplets, thus leading to an index of synergy of that variable R_Ω: The results are shown in Fig. 6, top left, where these indexes are compared with the first-order gradient as found by the proposed approach. Analogously, for each pair of variables we sum over the triplets containing that pair and obtain a synergy index and a redundancy index for all pairs of variables, depicted in Fig. 6, top middle, and compared with second-order gradients. In Fig. 6, top right, the distribution of Ω for all the significant triplets.

In the second row of Fig. 6, the same quantities are calculated using quadruplets: We stress that no synergistic quadruplet is found to be statistically significant.

These results show that, as far as the redundancy is concerned, the proposed approach leads to a pruning of redundant pairs of variables with respect to the index R_Ω, as shown by the vertical cloud of red points around $\partial_{ij}^2\Omega = 0$. Moreover, the redundant pattern of R_Ω is quite stable going from triplets to quadruplets. On the other hand, as far as the synergy is concerned, $\partial_i\Omega$ seems unrelated to S_Ω calculated on triplets.

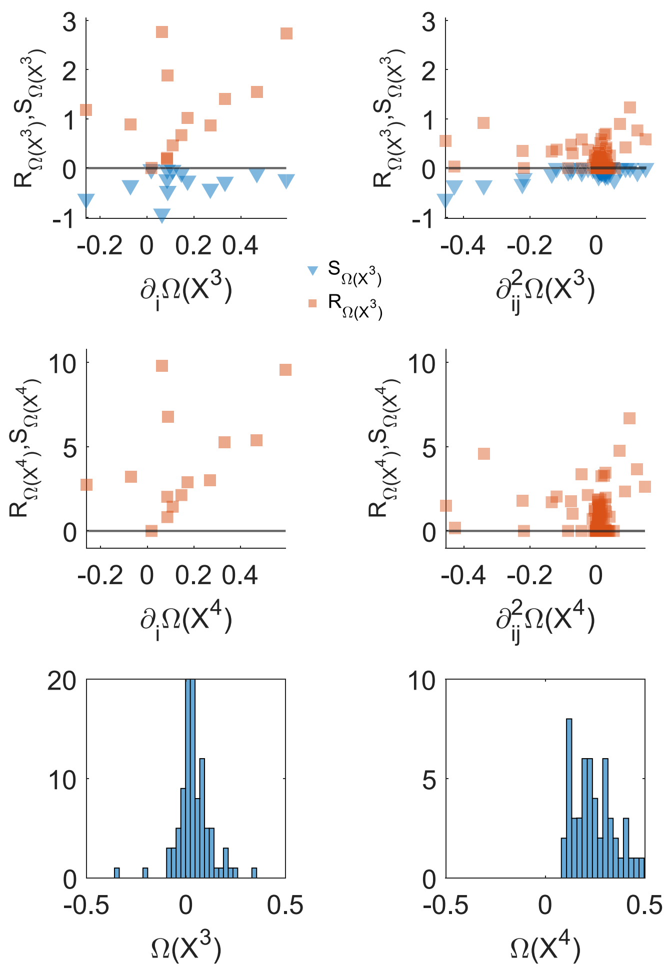

FIG. 6. The first row depicts the redundancy index R_Ω and the synergy index S_Ω in the univariate (left) and pairwise (right) case, see the text, plotted as a function the first order gradients and second order gradients, respectively, of the corresponding variable or pair of variables; in the second row the same analysis has been shown for the significant O-information quadruplets. Red and blue dots indicates R_Ω and S_Ω, respectively. In the third row we depict the distribution of the O-information values of all the significant triplets.

[1] A.-L. Barabasi and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, 2016).

[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. **424**, 175 (2006).

[3] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Networks beyond pairwise interactions: Structure and dynamics, Phys. Rep. **874**, 1 (2020).

[4] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, M. Murray, T. Peixoto, F. Vaccarino, and G. Petri, The physics of higher-order interactions in complex systems, Nat. Phys. **17**, 1093 (2021).

[5] F. Rosas, P. Mediano, A. Luppi, T. Varley, J. Lizier, S. Stramaglia, H. Jensen, and D. Marinazzo, Disentangling high-order mechanisms and high-order behaviours in complex systems, Nat. Phys. **18**, 476 (2022).

[6] P. L. Williams and R. D. Beer, Nonnegative decomposition of multivariate information, arXiv:1004.2515.

[7] J. T. Lizier, N. Bertschinger, J. Jost, and M. Wibral, Information decomposition of target effects from multi-source interactions:

Perspectives on previous, current and future work, Entropy **20**, 307 (2018).

[8] M. Wibral, V. Priesemann, J. W. Kay, J. T. Lizier, and W. A. Phillips, Partial information decomposition as a unified approach to the specification of neural goal functions, Brain and Cognition **112**, 25 (2017).

[9] A. Luppi, P. Mediano, F. Rosas, N. Holland, T. Fryer, J. O'Brien, J. Rowe, D. Menon, D. Bor, and E. Stamatakis, A synergistic core for human brain evolution and cognition, Nat. Neurosci. **25**, 771 (2022).

[10] O. Sporns, The complex brain: Connectivity, dynamics, information, Trends Cognit. Sci. **22**, 1185 (2022).

[11] F. E. Rosas, P. A. Mediano, M. Gastpar, and H. J. Jensen, Quantifying high-order interdependencies via multivariate extensions of the mutual information, Phys. Rev. E **100**, 032305 (2019) .

[12] M. Gatica, R. Cofré, P. A. Mediano, F. E. Rosas, P. Orio, I. Diez, S. P. Swinnen, and J. M. Cortes, High-order interdependencies in the aging brain, Brain Connectivity **11**, 734 (2021).

[13] S. Stramaglia, T. Scagliarini, B. C. Daniels, and D. Marinazzo, Quantifying dynamical high-order interdependencies from the O-information: An application to neural spiking dynamics, Front. Physiol. **11**, 1784 (2021).

[14] E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[15] Correspondingly, $\boldsymbol{X}^n_{-ik}$ indicates all the variables in $\boldsymbol{X}^n$ except X_i and X_k.

[16] W. McGill, Multivariate information transmission, Psychometrika **19**, 97 (1954).

[17] P. Baudot, M. Tapia, D. Bennequin, and J. M. Goaillard, Topological information data analysis, Entropy **21**, 869 (2019).

[18] T. Scagliarini, D. Marinazzo, Y. Guo, S. Stramaglia, and F. E. Rosas, Quantifying high-order interdependencies on individual patterns via the local o-information: Theory and applications to music analysis, Phys. Rev. Res. **4**, 013184 (2022).

[19] D. Marinazzo, L. Angelini, M. Pellicoro, and S. Stramaglia, Synergy as a warning sign of transitions: The case of the two-dimensional Ising model, Phys. Rev. E **99**, 040101(R) (2019).

[20] R. J. Baxter and F. Y. Wu, Exact Solution of an Ising Model with Three-Spin Interactions on a Triangular Lattice, Phys. Rev. Lett. **31**, 1294 (1973).

[21] L. Faes, G. Mijatovic, Y. Antonacci, R. Pernice, C. Barà, L. Sparacino, M. Sammartino, A. Porta, D. Marinazzo, and S. Stramaglia, A framework for the time- and frequency-domain assessment of high-order interactions in brain and physiological networks, IEEE Transact. Signal Proc. **70**, 5766 (2022).

[22] D. MacKinnon and S. Lamp, A unification of mediator, confounder, and collider effects, Prev. Sci. **22**, 1185 (2021).

[23] FRED, Federal reserve economic dataset, https://fred.stlouisfed.org/ (accessed May 5, 2022).

[24] F. Smets and R. Wouters, Shocks and frictions in U.S. business cycles: A Bayesian DSGE approach, in *European Central Bank, Working Paper Series* (ECB Publications, Frankfurt am Main, Germany, 2007), Vol. 722.

[25] A. Justiniano, G. E. Primiceri, and A. Tambalotti, Investment shocks and business cycles, J. Monetary Econ. **57**, 132 (2010).

[26] R. A. Ince, B. L. Giordano, C. Kayser, G. A. Rousselet, J. Gross, and P. G. Schyns, A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula, Human Brain Mapp. **38**, 1541 (2017).

[27] D. Marinazzo, J. Van Roozendaal, F. E. Rosas, M. Stella, R. Comolatti, N. Colenbier, S. Stramaglia, and Y. Rosseel, An information-theoretic approach to hypergraph psychometrics, arXiv:2205.01035.