

The low-rank hypothesis of complex systems

Received: 18 October 2022

Accepted: 24 October 2023

Published online: 10 January 2024

 Check for updates

Vincent Thibeault^{1,2}✉, Antoine Allard^{1,2} & Patrick Desrosiers^{1,2,3}✉

Complex systems are high-dimensional nonlinear dynamical systems with heterogeneous interactions among their constituents. To make interpretable predictions about their large-scale behaviour, it is typically assumed that these dynamics can be reduced to a few equations involving a low-rank matrix describing the network of interactions. Our Article sheds light on this low-rank hypothesis and questions its validity. Using fundamental theorems on singular-value decomposition, we probe the hypothesis for various random graphs, either by making explicit their low-rank formulation or by demonstrating the exponential decrease of their singular values. We verify the hypothesis for real networks by revealing the rapid decrease of their singular values, which has major consequences for their effective ranks. We then evaluate the impact of the low-rank hypothesis for general dynamical systems on networks through an optimal dimension reduction. This allows us to prove that recurrent neural networks can be exactly reduced, and we can connect the rapidly decreasing singular values of real networks to the dimension reduction error of the nonlinear dynamics they support. Finally, we prove that higher-order interactions naturally emerge from the dimension reduction, thus providing insights into the origin of higher-order interactions in complex systems.

Unravelling the emergent phenomena that drive the functions of complex systems requires us to bridge microscopic mechanisms with macroscopic ones. Rather than decomposing complex systems into as many components as possible, dimension reduction seeks a reduced system of macrostates or observables with a small enough dimension to get an insightful description but large enough to preserve the phenomena of interest. Yet, complex systems are characterized by extremely high dimensions—perhaps some sort of curse of dimensionality^{1–3}—and finding such reduced systems remains a challenge in several scientific disciplines.

In the paradigm ‘More is different’^{4,5}, it could appear contradictory to look for simple representations of complex systems. But ‘simple model’ does not mean ‘simple behaviour’. The logistic equation⁶, cellular automata^{7,8} and spin glasses^{9,10} exhibit complex behaviours such as chaos, and recurrent neural networks (RNNs) can approximate any finite trajectory of N -dimensional dynamical systems¹¹.

In network science, the topology of the interactions among the constituents of complex systems is typically simplified to a graph,

defined by a set of vertices and a set of edges (Fig. 1a,b). Such a representation allows us to extract the dominant properties of complex networks, such as their organization into modules¹². An ongoing change of paradigm is to use hypergraphs or simplicial complexes rather than graphs to take into account the higher-order interactions observed in some real-world systems^{13,14}. In addition to finding an appropriate dimension to describe a complex system, one has to uncover the orders of its interactions. As shown later, both problems are intertwined.

A graph can always be described as a matrix. This simple, yet essential, possibility unlocks several tools from linear algebra that can be used to characterize networks. Among them, spectral theory can identify the fundamental components of a matrix through matrix decomposition. Eigenvalue decomposition has long been used to extract key properties of graphs, such as their invariants¹⁵, their modular structure¹⁶, the centrality of their vertices¹⁷ or the bifurcations of a dynamical system represented by a network¹⁸.

¹Département de physique, de génie physique et d’optique, Université Laval, Québec, Québec, Canada. ²Centre interdisciplinaire en modélisation mathématique de l’Université Laval, Québec, Québec, Canada. ³Centre de recherche CERVO, Québec, Québec, Canada.

✉ e-mail: vincent.thibeault@ulaval.ca; patrick.desrosiers@phy.ulaval.ca

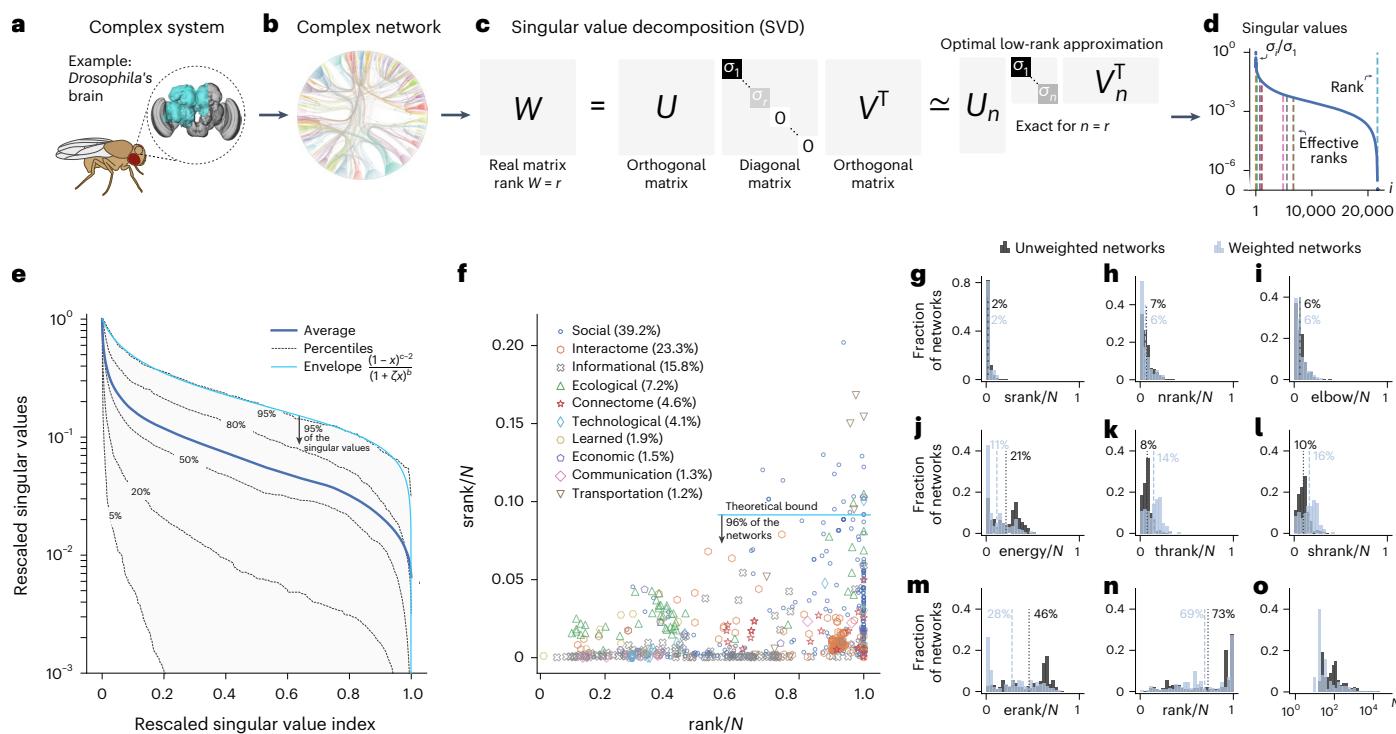


Fig. 1 | Experimental verification of the low-rank hypothesis for real networks.

a, *Drosophila melanogaster*'s hemibrain is an example of a complex system. The open-source image of the hemibrain is from ref. 69. **b**, A complex network illustration of *Drosophila melanogaster*'s connectome⁶⁹. Of the 21,733 vertices, 5% were randomly selected for the visualization. **c**, SVD of a real matrix of rank r . The truncated SVD is the optimal low-rank approximation of the matrix, as guaranteed by the Schmidt–Eckart–Young–Mirsky theorem (Supplementary Theorem 8). **d**, Rapid decrease of the singular values of the matrix describing the *Drosophila melanogaster*'s connectome with the ordinates in logarithmic scale. The vertical dashed lines indicate the rank of the matrix as well as seven measures of effective rank (Table 2). **e**, The average and percentiles of the singular value distribution of 679 real networks of different origins rescaled by their respective largest singular value (Methods). The shaded background is the

region between the 5th and the 95th percentiles. The parameters of the singular-value (hypergeometric) envelope above 95% of all the singular values are $b \approx 0.54$, $c \approx 2.3$ and $\zeta \approx 25$. **f**, The stable rank to dimension ratio versus the rank to dimension ratio for real networks. 96% of the networks have a stable rank below the theoretical bound, which is obtained from the singular-value envelope in **e** and Theorem 3 (Methods). The approximate percentage of networks in each category is given in parentheses beside the name of the category. **g–m**, Fraction of the 679 real networks (502 unweighted networks and 177 weighted networks) versus effective rank divided by N : versus srank (**g**), versus nrk (**h**), versus elbow (**i**), versus energy (**j**), versus thrank (**k**), versus shrank (**l**) and versus erank (**m**). **n**, Fraction of the networks versus rank divided by N . **o**, Fraction of the networks versus number of vertices N shown in log scale. In **g–o**, the vertical dashed lines labelled with a percentage are the averages for the distributions.

One pressing challenge in network science is to efficiently adapt the tools of spectral theory to directed, weighted and signed (for example, excitatory-inhibitory) networks and, hence, to general real matrices. Indeed, eigenvalue decomposition yields complex eigenvalues and complex-valued eigenvectors in general, potentially causing methodological problems (sections IID and IIF in the Supplementary Information). Worse still, it is not even guaranteed that the matrix representation of a network is diagonalizable. For instance, neither the trivial directed graph with two vertices connected by one directed edge nor any network whose (real) matrix representation W is rectangular is diagonalizable (for example, an incidence matrix or an interlayer matrix in multilayer networks).

Yet, the matrices WW^\top and $W^\top W$ are always square, symmetric and, thus, diagonalizable, which lays the foundations of singular-value decomposition (SVD; Fig. 1c and Supplementary Theorem 1). Interestingly, the decomposition exists for any matrix, the singular vectors are real-valued, and the singular values $\sigma_1, \dots, \sigma_N$ are non-negative real numbers. Notably, the number of non-zero singular values equals the rank of W . Moreover, SVD inherits various theorems from eigenvalue decomposition¹⁹, such as Weyl's theorem^{20,21}, but it also produces new fundamental results. In particular, SVD is a central tool for dimension reduction in general, partly because the Schmidt–Eckart–Young–Mirsky theorem guarantees that the

truncated SVD yields the best low-rank approximation of a matrix (Fig. 1c and Supplementary Theorem 8).

The salient properties of SVD and its close relationship with the (effective) rank of a matrix have not yet been completely recognized in network science and spectral graph theory, compared to its ubiquity in data science (for example, matrix completion²², dynamic mode decomposition²³ and optimal singular-value shrinkage²⁴), control theory (for example, the Kalman criterion^{25–27}), random matrix theory (for example, Marčenko–Pastur's law²⁸) and linear algebra (for example, matrix norms¹⁹). SVD is not even mentioned in many of the main introductory textbooks of network science or spectral graph theory (Supplementary Information section IIA).

Throughout the Article, we leverage the key attributes of SVD to define and evaluate the impact of the low-rank hypothesis of complex systems. Before tackling complex systems as high-dimensional nonlinear dynamical systems, we first describe the theoretical evidence supporting the hypothesis for random graphs, which is followed by an empirical verification of the hypothesis for real networks.

Evidence supporting the hypothesis for network models

It is first instructive to consider random graphs, that is, sets of graphs equipped with a probability measure that depends on some properties,

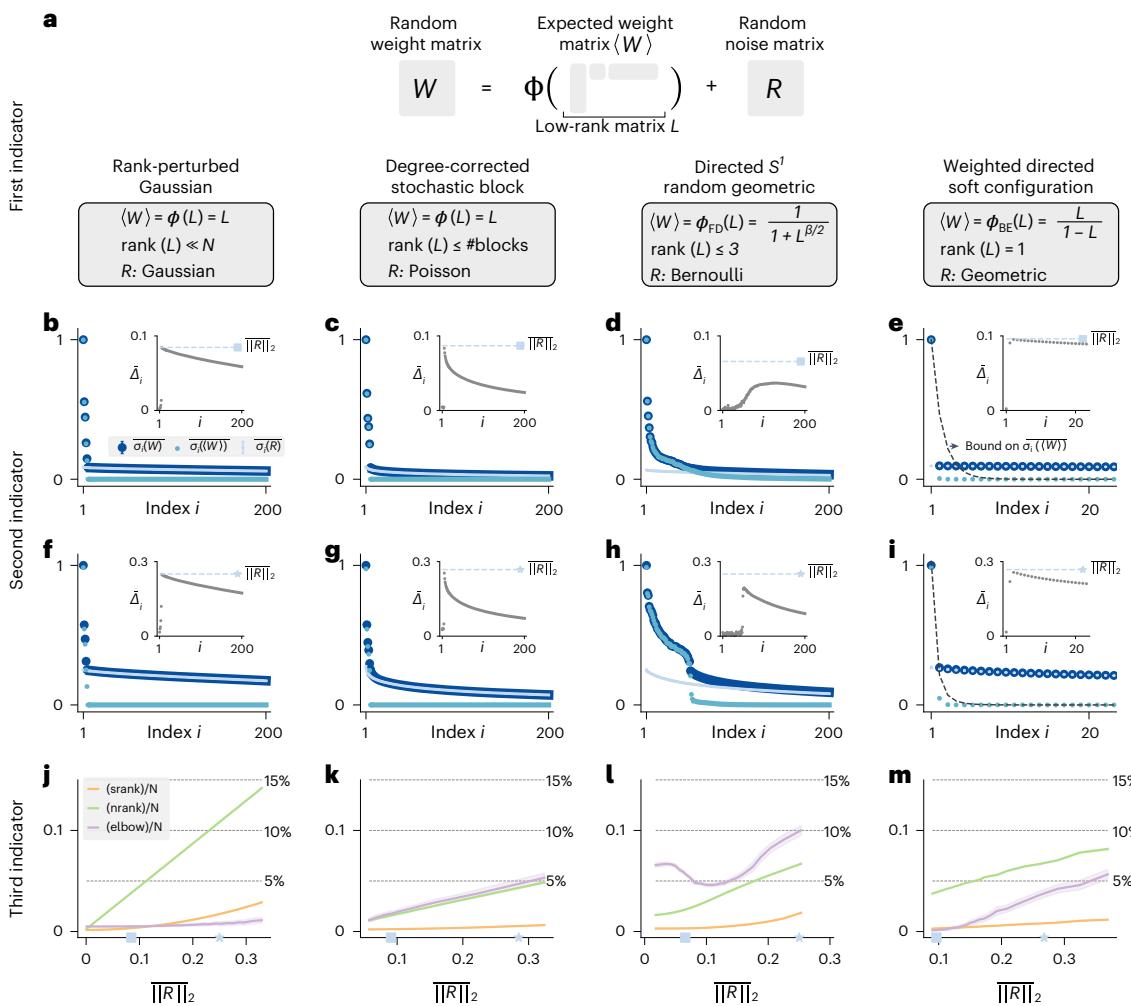


Fig. 2 | Three indicators of the low-rank hypothesis for random graphs.

a, Many random graphs have a random matrix representation in which the expected weight matrix $\langle W \rangle$ is a matrix-valued function Φ of a low-rank matrix L plus a centred random part R . Four examples of random matrices with different weight distributions and functions Φ (the rank-perturbed Gaussian model (RPG), the degree-corrected stochastic block model (DCSBM), the directed S^1 random geometric model (S^1 RGM) and the weighted directed soft configuration model (WDSCM)) are illustrated and aligned with the corresponding panels below. The functions ϕ_{FD} and ϕ_{BE} , respectively, stand for a Fermi–Dirac distribution with inverse temperature β and a Bose–Einstein distribution in which the division is element-wise, for example, the element (i,j) of $L/(1-L)$ is $L_{ij}/(1-L_{ij})$. **b–i**, Rescaled and averaged singular values of the random weight matrix, its expected part and its random part for each random graph (RPG **b** and **f**, DCSBM **c** and **g**, S^1 RGM **d** and **h**, WDSCM **e** and **i**) are shown in two noise regimes: square

markers for $\|R\|_2$ near 0.1 (**b–e**) and star markers for $\|R\|_2$ near 0.3 (**f–i**). The singular values are respectively denoted $\sigma_i(W)$, $\bar{\sigma}_i(W)$ and $\sigma_i(R)$ (from darker to lighter blue markers) where $\bar{x} = \langle x \rangle / (\|W\|_2)$ and $\langle \cdot \rangle$ denotes the average over the ensemble of graphs. Error bars indicate the standard deviation of the singular values but are too small to be seen. The random graphs have $N = 10^3$ vertices and only the first 200 (or 20 in **e** and **i**) singular values are shown for clarity. The dashed black lines in **e** and **i** are the rescaled upper bounds on the singular values of $\langle W \rangle$ in Theorem 1 (Methods) with RMSEs over all $i \in \{1, \dots, N\}$ of 0.02 in **e** and 0.006 in **i**. The insets show the rescaled and averaged Δ_i and its upper bound defined in equation (1). **j–m**, Evolution of three effective ranks (srank, nrank and thrank, averaged over the ensemble of graphs and rescaled by N) according to the strength of the noise $\|R\|_2$ for RPG (**j**), DCSBM (**k**), S^1 RGM (**l**) and WDSCM (**m**). The shaded areas are the standard deviations of the effective ranks. The parameter values used for each random graph can be found in Methods.

such as the degrees, the modules or the distance between vertices in some metric space (Supplementary Information sections IIA and IIB). Mathematically, they can always be written as random matrices $W = \langle W \rangle + R$, where $\langle W \rangle$ is the expected weight matrix and R is a random matrix with mean 0.

By examining many widely used random graphs, we observed that their expected matrices involve low-rank matrices. Indeed, we highlight the—usually implicit—assumption that $\langle W \rangle$ is equal to a function Φ of a low-rank matrix L (Fig. 2a, Table 1 and Supplementary Information section IIA). In many cases, $\Phi(L) = L$ and it is straightforward to see the low rank of $\langle W \rangle$ since it can be written into its rank-factorized form. A particular Weyl inequality already establishes an expected, but important, outcome of the hypothesis: a small random part R ensures that each singular value of W is close to the relevant value in $\langle W \rangle$, that is,

$$\Delta_i = |\sigma_i(W) - \sigma_i(\langle W \rangle)| \leq \|R\|_2, \quad (1)$$

for all $i \in \{1, \dots, N\}$, where $\sigma_i(A)$ denotes the i th singular value of A and $\|\cdot\|_2$ denotes the spectral matrix norm (Supplementary Theorem 5 and Supplementary Corollary 7). Viewing $W = \langle W \rangle + R$ with $\langle W \rangle = L$ and $\text{rank}(L) = r$ as a spiked random matrix^{29–33} offers an even more precise perspective. For such matrices, the singular values have a ‘bulk’ related to the singular values of R and the creation or annihilation of outlying singular values is asymptotically characterized by the Baik–Ben Arous–Péché phase transition³⁴. Notably, the presence of d singular-value outliers for W depends only on a threshold on the dominant singular values of $\langle W \rangle$, namely $\sigma_1(\langle W \rangle), \dots, \sigma_d(\langle W \rangle)$ (ref. 32; Supplementary Information section IIA). Therefore, a low rank r for $\langle W \rangle$ together with mild threshold conditions imply that the largest singular values of W

Table 1 | Low-rank matrix L characterizing the expected adjacency matrix for different random graphs of N vertices

| Model | Low-rank matrix L | rank(L) | $\phi(L_{ij})$ |
|------------|---|--|----------------|
| Unweighted | $Np \hat{\mathbf{1}}\hat{\mathbf{1}}^\top$ | 1 | L_{ij} |
| | $\frac{\ \kappa\ ^2}{2M} \tilde{\mathbf{K}}\tilde{\mathbf{K}}^\top$ | 1 | L_{ij} |
| | DSCM | $\ \alpha\ \ \beta\ \hat{\mathbf{a}}\hat{\beta}^\top$ | 1 |
| | MD | $\sum_{\mu,\nu=1}^r \Delta_{\mu\nu} \mathbf{v}_\mu \mathbf{v}_\nu^\top$ | r |
| | SBM | $\sum_{\mu,\nu=1}^q \sqrt{n_\mu n_\nu} p_{\mu\nu} \mathbf{b}_\mu \mathbf{b}_\nu^\top$ | $\leq q$ |
| | $S^D \text{RGM}$ | $\frac{R^2}{\mu^2} (\mathbf{K}_{\text{in}} \tilde{\mathbf{K}}_{\text{out}}^\top) \circ \bar{\theta}$ | $\leq D+2$ |
| Weighted | $\mathcal{G}(N, p, w)$ | $Npw \hat{\mathbf{1}}\hat{\mathbf{1}}^\top$ | 1 |
| | WCL | $\mathbf{y}\mathbf{y}^\top$ | 1 |
| | WDSCM | $\mathbf{y}\bar{\mathbf{y}}^\top$ | 1 |
| | RPG | $\sum_{\mu=1}^r \mathbf{m}_\mu \mathbf{n}_\mu^\top$ | r |
| | WSBM | $\sum_{\kappa,\nu=1}^q \sqrt{n_\kappa n_\nu} \mu_{\kappa\nu} \mathbf{b}_\kappa \mathbf{b}_\nu^\top$ | $\leq q$ |
| | DCSBM | $\Lambda \circ (\mathbf{K}_{\text{in}} \tilde{\mathbf{K}}_{\text{out}}^\top)$ | $\leq q$ |
| | RDPG | $\sum_{\mu=1}^d \mathbf{X}_\mu \mathbf{X}_\mu^\top$ | $\leq d$ |

CL, Chung–Lu; DCSBM, degree-corrected stochastic block model; DSCM, directed soft configuration model; MD, metadegree; RDPG, random dot product graph; RGM, random geometric model; RPG, rank-perturbed Gaussian; SBM, stochastic block model. ‘W’ in front of an acronym stands for ‘weighted’. For S^D RGM, the rank of L is, more precisely, $D, D+1$ or $D+2$, as a consequence of Theorem 7 in ref. 66 and the inequality $\text{rank}(A \circ B) \leq \text{rank}(A) \text{rank}(B)$. The parameters q, r, d and D are usually assumed to be small compared to N . More details about these random graphs and others are given in Supplementary Information section II A.

are in the vicinity of $\sigma_1(\langle W \rangle), \dots, \sigma_r(\langle W \rangle)$, which is the first indicator of the low-rank hypothesis.

However, the low rank of $\langle W \rangle$ is not always obvious, such as for DSCM and its weighted version. Indeed, their expected weight matrices are nonlinear functions of rank-one matrices (Methods). Leveraging Weyl’s inequalities, we demonstrated for both models that the singular values of $\langle W \rangle$ are bounded above by an exponentially decreasing term (Theorem 1 (Methods) and Fig. 2e,i). Figure 2b–i illustrates how the singular values of W for four different weighted random graphs and two noise regimes inherit the decreasing trend of the dominant singular values of $\langle W \rangle$, whereas the subdominant ones are related to R . The rapid decrease of the dominant singular values of W hints at the approximate low rank of the network and, thus, constitutes a second crucial indicator of the low-rank hypothesis.

The attributes ‘rapid decrease’ and ‘approximate low rank’ remain to be quantified, however. To do so, we invoke effective ranks. For instance, the stable rank measures the relative importance of the squared singular values with respect to σ_1^2 (Table 2). In Fig. 2j–m, we depict its persistence with an increase of the noise level for four random graphs. How ‘low’ is an effective rank of a random graph is better understood through its asymptotic behaviour as $N \rightarrow \infty$ (Methods). Different singular value decreases lead to different asymptotic behaviours for the effective ranks, from constant $O(1)$ and sublinear growth $O(N^{1-\epsilon})$ with $\epsilon \in (0, 1]$ to linear growth $O(N)$ (Supplementary Information section IIC). Notably, sublinear growth implies that the effective ranks to dimension ratio falls to zero asymptotically as $O(N^{-\epsilon})$. We will thus say that an effective rank is low if it grows at most sublinearly. For example, we demonstrate that for any growing network model with exponentially decreasing singular values (for example, soft configuration models), the lowest asymptotic behaviour $O(1)$ for the stable rank and two other effective ranks (Corollary 2) are implied. However, when dealing with

Table 2 | Different effective ranks of a matrix of dimension $N \times N$ and of rank r expressed in terms of its singular values $\sigma_1 \geq \dots \geq \sigma_N$

| Abbreviation | Expression |
|--------------|---|
| srank | $\sum_{i=1}^r \sigma_i^2 / \sigma_1^2$ |
| nrank | $\sum_{i=1}^r \sigma_i / \sigma_1$ |
| energy | $\min \left[\arg \max_{\ell \in \{1, \dots, N\}} \left(\sum_{i=1}^\ell \sigma_i^2 / \sum_{j=1}^r \sigma_j^2 \right) \right]$ |
| elbow | $\frac{1}{\sqrt{2}} \arg \max_{i \in \{1, \dots, N\}} \left \frac{i-1}{N-1} + \frac{\sigma_i - \sigma_N}{\sigma_1 - \sigma_N} - 1 \right - 1$ |
| erank | $\exp \left[- \sum_{i=1}^r \frac{\sigma_i}{\sum_{j=1}^r \sigma_j} \log \frac{\sigma_i}{\sum_{j=1}^r \sigma_j} \right]$ |
| thrark | $\#\left\{ \sigma_i \mid i \in \{1, \dots, N\} \text{ and } \sigma_i > \frac{4\sigma_{\text{med}}}{\sqrt{3}\mu_{\text{med}}} \right\}$ |
| shrank | $\#\{s^*(\sigma_i) \mid i \in \{1, \dots, N\} \text{ and } s^*(\sigma_i) > 0\}$ |

For energy, the constant τ is a threshold to be set between 0 and 1. For thrark, σ_{med} is the median singular value and μ_{med} is the median of a Marčenko–Pastur probability density function⁶⁷. For shrank, s^* denotes an optimal singular-value shrinkage function^{24,68}. The complete names and the details of each of the effective ranks are given in Supplementary Information section IC.

a single instance of a random graph or with a real network, N should be kept fixed and the above asymptotic perspective is not applicable. Yet, we can give a more subtle, graded response to the question ‘how low?’ with effective rank to dimension ratios: values much smaller than 1 indicate that few singular values contribute appreciably in the SVD, meaning that W can be well approximated by a low-rank matrix. Having small effective rank to dimension ratios is thus a third indicator, this time quantitative, supporting the low-rank hypothesis.

In summary, the low-rank hypothesis has been described with three indicators for random graphs. The second one, the rapid decrease of the singular values, is the central indicator of the hypothesis: the first indicator being a theoretical cause for the decrease and the third indicator being a consequence. The second and third indicators are not tied to any theoretical model and can be applied to any type of networked data. We, hence, adopt the following general, yet workable, definition of the low-rank hypothesis: it is the assumption that the singular values of a network’s weight matrix decrease rapidly, implying low effective ranks. We now put this hypothesis to the test.

Verification of the hypothesis for real networks

Despite its frequent use—often implicit, but sometimes very explicit^{35,36}—the low-rank hypothesis has yet to be verified experimentally for real networks in all their diversity.

Our experiments revealed that the rapid decay of the singular values in real networks is the norm. As an example, we illustrate the singular-value profile of the connectome of *Drosophila melanogaster* in Fig. 1d. Figure 1e presents a coalesced view of the singular-value profiles for 679 real networks from ten different origins. As a guide to appreciate the decreases, we trace a general singular-value envelope below which 95% of the singular values of all the networks belong.

Having an explicit form for the singular-value envelope allows us to interpret the stable rank as the area under a curve (Supplementary Information section IIC) and then to find a theoretical bound below which most of the stable ranks of the networks lie (Theorem 3 (Methods)). In Fig. 1f, we illustrate the stable rank of the real networks along with the theoretical bound below which there are 96% of the networks, indicating that the stable rank is generally expected to be less than 10% of the number of vertices N .

To ensure that this observation is not limited to the stable rank, we report in Fig. 1g–m similar observations for other effective ranks (Methods). It is not surprising that nrank and erank are larger than srank. In fact, it is easily shown that $\text{srank} \leq \text{nrank} \leq \text{erank} \leq \text{rank}$

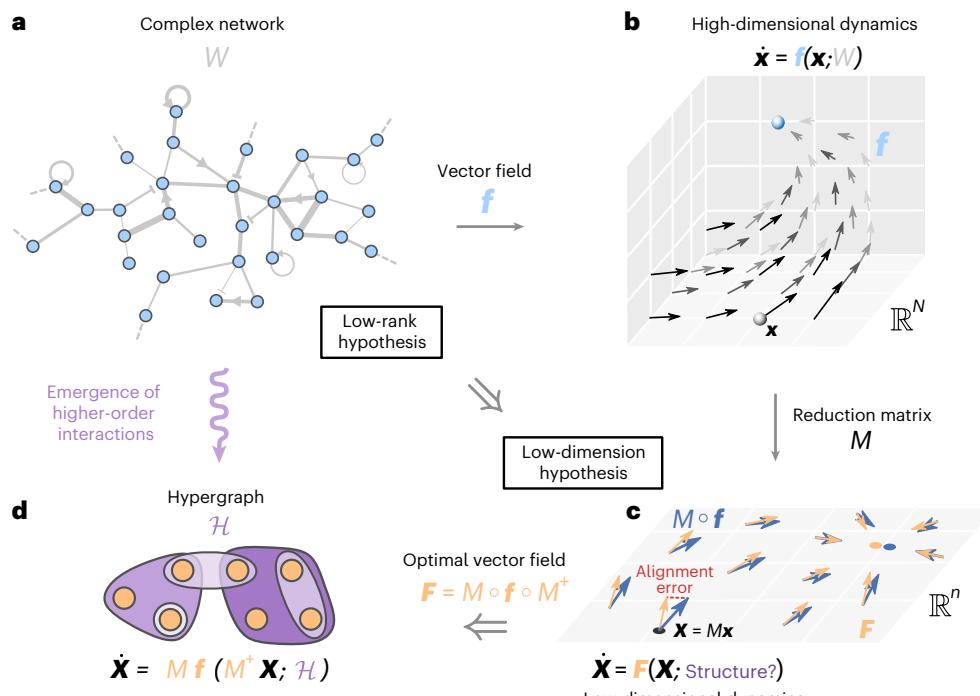


Fig. 3 | The low-rank hypothesis of complex systems and the emergence of higher-order interactions. **a**, A complex network represented as a weighted (edge widths), signed and directed (edges with arrows or a perpendicular line for inhibition) graph with weight matrix W . **b**, A vector field \mathbf{f} of an N -dimensional dynamical system on a network converging to an equilibrium point. **c**, Dimension reduction of a dynamical system through the reduction matrix, which is a linear transformation $M : \mathbb{R}^N \rightarrow \mathbb{R}^n$, so that $\mathbf{x} \mapsto \mathbf{X} = M\mathbf{x}$. The blue arrows illustrate the exact vector field $M \circ \mathbf{f}$ in \mathbb{R}^n (where \circ is the element-wise product) whereas the

orange arrows represent an approximate vector field \mathbf{F} . Dimension reduction is about aligning the vector fields, that is, minimizing alignment errors. **d**, The least-square optimal vector field $M \circ \mathbf{f} \circ M^+$ yields higher-order interactions between the observables X_1, \dots, X_n represented by some general hypergraph \mathcal{H} with n vertices. The hyperedges are represented by the shaded regions. Their weights and orientations (Supplementary Information section III C) are not illustrated to avoid cluttering the figure. Note that we make a slight abuse of notation by considering \mathbf{x} (respectively \mathbf{X}) as a function of time and also as a point in \mathbb{R}^N (respectively \mathbb{R}^n).

(Methods). Unlike the effective ranks, the rank of real networks is often comparable to their dimension (Fig. 1*n*). This observation is expected, especially for weighted networks with real weights, since non-invertible matrices form a set of measure 0.

The datasets considered are for real networks with fixed N , but the asymptotic behaviours of their effective ranks can still be evaluated as if there were a related growing graph whose singular values remain within experimental singular-value envelopes as N grows. Using this approach, we prove that singular-value envelopes such as the one in Fig. 1*e* admits constant and sublinear growth for srank, nrank and erank (Methods).

Overall, we show that many real networks have rapidly decreasing singular values, leading to low effective ranks. Interestingly, such observations seem to be widespread for big data matrices^{37–39}, but they remain puzzling. In particular, the consequences of these observations for high-dimensional nonlinear dynamics on networks are still to be untangled, which is the subject of the next section.

Induced low-dimension hypothesis

Intuitively, we expect that having low (effective) rank networks will enable the dimension reduction of the dynamics for these networks. Consider the complete dynamics $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; W)$, where $\mathbf{x}(t) \in \mathbb{R}^N$ is the system's state at time t , $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a continuously differentiable vector field and W is an $N \times N$ weight matrix describing the network (Fig. 3a,b). More specifically, given $\mathbf{g} : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^N$ and $W(\mathbf{x}(t)$ is unknown), we examine the subclass of dynamics $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{y})$ where $\mathbf{y} = W\mathbf{x}$.

This subclass of dynamics already highlights an important implication of the low-rank hypothesis. The linear function $\mathbf{x} \mapsto \mathbf{y} = W\mathbf{x}$ in \mathbf{g} has a very special role: even if \mathbf{x} is part of an N -dimensional manifold,

when W has a low rank, the vector in the image of W will be part of a low-dimension submanifold. Even if W has full rank, our experimental observations in Fig. 1 show that it probably has a low effective rank. We can, hence, say that $W\mathbf{x}$ will be part of an effectively low-dimension submanifold.

Just as some random graph models are crafted from a nonlinear function ϕ of a low-rank matrix L (Fig. 2a), the vector field \mathbf{g} depends nonlinearly on $W\mathbf{x}$, making it challenging to assess the low dimensionality of $\mathbf{g}(\mathbf{x}, \mathbf{y})$. Despite recent developments^{40–46}, it remains unclear how to choose a dimension for the reduced dynamics and how to quantify the reduction error for nonlinear dynamics on complex networks.

The dimension reduction of dynamical systems can be imagined as the problem of aligning a low-dimensional vector field with its high-dimensional counterpart (Fig. 3c and Supplementary Information section III A). This involves selecting an $n \times N$ reduction matrix M that maps the elements of the complete system to the reduced system as well as a vector field \mathbf{F} describing the evolution of a set of observables $(X_\mu)_{\mu=1}^n$ in \mathbb{R}^n . The alignment error in \mathbb{R}^n at $\mathbf{x} \in \mathbb{R}^N$, denoted $\mathcal{E}(\mathbf{x})$, can then be defined as the error between the vector fields $M \circ \mathbf{f}$ and $\mathbf{F} \circ M$ (Methods).

Minimizing the alignment error to find the optimal pair (M, \mathbf{F}) is challenging in general (Supplementary Information section III A), and the best choice hinges on the modeller's objective. For instance, selecting M to ensure that the temporal evolution of \mathbf{X} remains interpretable throughout time (for example, synchronization observables⁴⁶), might further complicate the optimization problem.

Let us concentrate on identifying \mathbf{F} without taking into account M for now. Using least squares, we proved that $M \circ \mathbf{f} \circ M^+$ minimizes the alignment error in \mathbb{R}^N , where superscript + denotes pseudoinversion

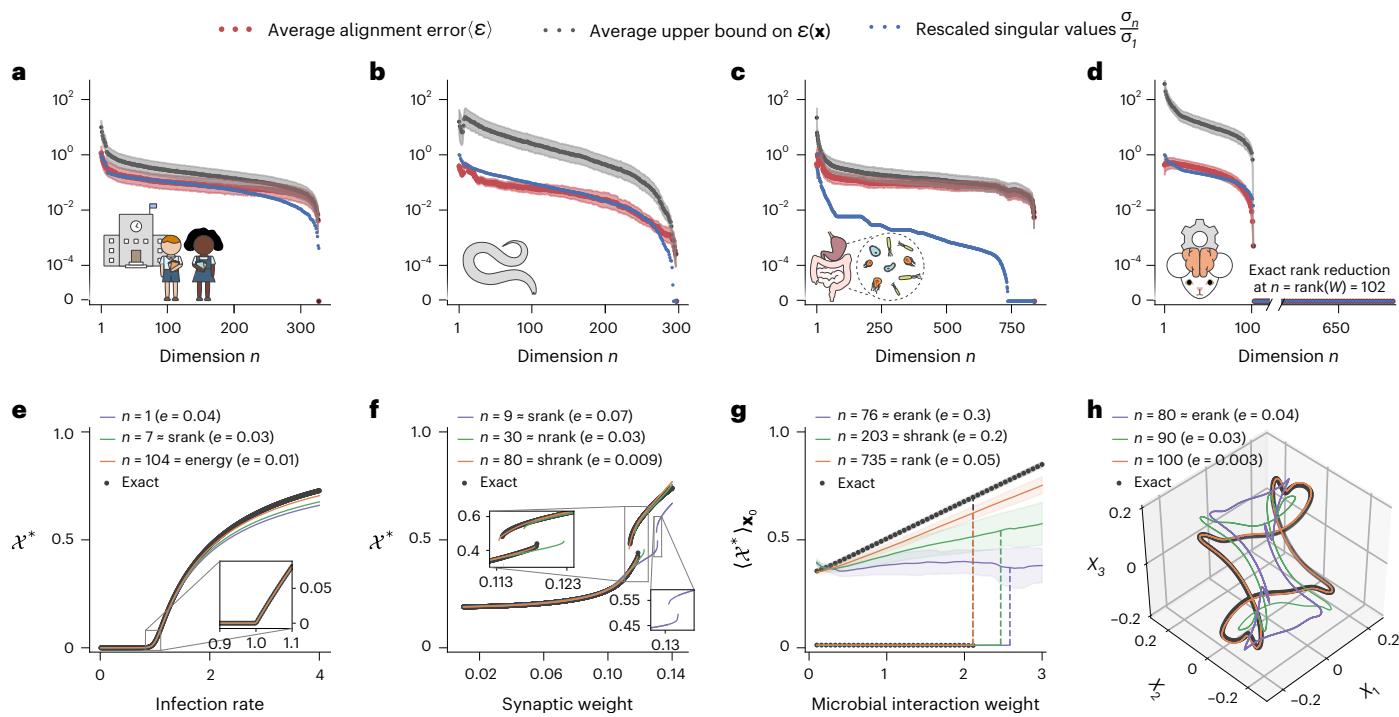


Fig. 4 | Dimension reduction errors for nonlinear dynamics on real complex networks related to their singular values and effective ranks. **a–d.** The decrease of the alignment error $\mathcal{E}(x)$ (red markers) is in accordance with the rapid decrease of singular values (blue markers), as expected by the analytical upper bound in equation (4) (solid black line) for an epidemiological network (**a**), a neuronal network (**b**), a microbial network (**c**) and an RNN (**d**). The shaded regions in grey and light red represent the standard deviation of the upper bound and the error, respectively. For every value of n , we have 10^3 different samples for x and for the respective parameters of each dynamics. The upper bounds are computed exactly in **a** and **c**, whereas they are approximated in **b** and **d** (see the details in Supplementary Information section IIIE). **e–h.** Comparison of the bifurcation diagrams (respectively trajectories in **h**) for the global observable, denoted $x^* = w \cdot x^*$ at equilibrium where w is an $n \times 1$ real vector specific to the dynamics, of the complete dynamics (black markers) versus the reduced dynamics (solid coloured lines) at different dimensions n . The RMSEs e are computed between the global equilibrium points of the complete and the reduced epidemiological

dynamics (**e**), neuronal dynamics (**f**), microbial dynamics (**g**), RNN dynamics (**h**) at different n (Methods). **a,e.** Epidemiological dynamics (quenched mean-field susceptible-infected-susceptible) on a high-school contact network ($N = 327$, undirected, binary) rescaled by the largest singular value. **b,f.** Neuronal (Wilson–Cowan) dynamics on the *C. elegans* connectome ($N = 297$, signed, weighted, directed). **c,g.** Microbial population dynamics on a human gut microbiome network ($N = 838$, signed, weighted, directed). Note that there are several stable upper branches depending on the initial condition x_0 (Methods). Here we show the upper branches averaged over x_0 (black markers and solid coloured lines) with the standard deviation (shaded regions) and we show one lower branch. The loss of stability of the lower branch is indicated by a dashed vertical line that connects it, for visualization, to the average of the upper branches. **d,h.** RNN dynamics on a learned network ($N = 669$, signed, weighted, directed) for which we have shrunk its singular values using optimal shrinkage with the Frobenius norm²⁸ to emphasize that dimension reduction for the RNN dynamics is exact when n is the rank of the network (Methods).

(Methods). Doing so allowed us to show, for $\dot{x} = g(x, y)$, that the alignment error $\mathcal{E}(x)$ caused by the least-square vector field satisfies

$$\sqrt{n} \mathcal{E}(x) \leq \|M'_x(I - M^+ M)x\| + \|W(I - M^+ M)\|_2 \|M'_y\|_2 \|x\|, \quad (2)$$

where J'_x and J'_y are Jacobian matrices (Methods).

Interestingly, the previous inequality suggests a non-arbitrary way of selecting the reduction matrix. Indeed,

$$M = V_n^\top \quad (3)$$

minimizes the factor $\|W(I - M^+ M)\|_2$ related to the interactions in the system, thus generally making each observable X_μ global, that is, it contains information about most vertices (Methods).

The choice made in equation (3) prompted us to derive another inequality revealing the contribution of the network's singular values to the alignment error (Theorem 4):

$$\sqrt{n} \mathcal{E}(x) \leq \|V_n^\top J'_x(I - P)x\| + \sigma_{n+1} \|V_n^\top J'_y\|_2 \|x\|, \quad (4)$$

where $P = V_n V_n^\top$. Notably, the inequality provides a criterion for exact dimension reduction: if $J'_x = dI$ for $d \in \mathbb{R}$ and $n = \text{rank}(W)$, the upper

bound vanishes to zero and the dimension reduction is exact (Methods). Consequently, a general class of dynamics, including RNNs and Wilson–Cowan neuronal dynamics, can be exactly reduced (Methods). The upper bound (4) is meant to be intuitive (not necessarily tight): it connects the swift decay of singular values of a network with the dimension reduction error. As a basic example, the relative alignment error $\mathcal{E}(x)/\|x\|$ for the linear system $\dot{x} = Wx$ is simply upper-bounded by σ_{n+1}/\sqrt{n} , meaning that a rapid decrease of the singular values of W , even if related to an arbitrarily weighted network, directly induces a rapid decrease of the alignment error.

Figure 4a–d illustrates the decrease of the alignment error with n —the latter being in accordance with the rapid decay of the upper bound and of the singular values—for the dynamics on four real networks. We show how n can be tuned to predict an epidemic from the epidemiological dynamics (Fig. 4e), a hysteresis from the neuronal dynamics (Fig. 4f), stable branches from the microbial dynamics (Fig. 4g) or a limit cycle in an RNN (Fig. 4h). Although effective ranks can help when selecting a suitable dimension n to describe a collective phenomenon, we use them only as an indication: n should be chosen according to the modeller's tolerance to qualitative (for example, is the hysteresis preserved?) or quantitative (for example, is the predicted transition accurate?) errors. Thus, it becomes clear that having low

(effective) rank matrices describing complex networks enables the dimension reduction of the nonlinear dynamics on these networks.

The reduced system is akin to the low-dimensional dynamics on a smaller structure whose nature remains to be specified (Fig. 3c). We show in the next section that dimension reduction ultimately leads to the emergence of higher-order interactions, as illustrated in Fig. 3d.

Emergence of higher-order interactions

Theoretical and experimental evidence for the existence of higher-order interactions in various complex systems has been reported, and its consequences—for example, on explosive transitions⁴⁷ or mesoscopic localization⁴⁸—have been extensively studied⁴⁹. However, the origin of these interactions remains under active investigation, notably for oscillatory systems^{50,51} (Supplementary Information section IIIC).

Using our framework, a simple example readily provides insights into the emergence of higher-order interactions. Consider the epidemiological dynamics $\dot{x}_i = -d_i x_i + \gamma(1 - x_i)y_i$ with $i \in \{1, \dots, N\}$, where x_i is the probability that the vertex i is infected, $y = Wx$ whereas d_i and γ denote the recovery rate of vertex i and the infection rate, respectively. The reduced dynamics is then given by

$$\dot{X}_\mu = \sum_{v=1}^n (\mathcal{D}_{\mu v} + \mathcal{W}_{\mu v}) X_v - \gamma \sum_{i=1}^N M_{\mu i} \left(\sum_{v=1}^n M_{iv}^+ X_v \right) \left(\sum_{j=1}^N \sum_{k=1}^n W_{ij} M_{jk}^+ X_k \right) \quad (5)$$

for all $\mu \in \{1, \dots, n\}$, where $\mathcal{D} = -MDM^+$ is a reduced $n \times n$ recovery rate matrix with $D = \text{diag}(d_1, \dots, d_N)$, and $\mathcal{W} = \gamma MWM^+$ is a reduced $n \times n$ weight matrix.

Let us inspect the last term in equation (5) more carefully. For simplicity, consider that $M^+ = M^\top$, that is, M has orthogonal rows. Then, $M_{\mu i}$ quantifies the influence of vertex i on the μ th observable, $M_{\mu i}^\top X_v$ is the influence of the v th observable weighted by its dependence over vertex i and $W_{ij} M_{jk}^\top X_k$ is the influence of the k th observable weighted by its dependence over vertex j that connects to vertex i . Altogether, these factors form a third-order interaction between the observables X_μ , X_v and X_k , an observation that is made more explicit by rearranging equation (5) as

$$\dot{X}_\mu = \sum_{v=1}^n (\mathcal{D}_{\mu v} + \mathcal{W}_{\mu v}) X_v + \sum_{v,k=1}^n \mathcal{T}_{\mu v k} X_v X_k, \quad (6)$$

where the third-order interactions are encoded in a third-order tensor \mathcal{T} with elements

$$\mathcal{T}_{\mu v k} = -\gamma \sum_{i,j=1}^N M_{\mu i} M_{iv}^+ W_{ij} M_{jk}^+ \quad (7)$$

for $\mu, v, k \in \{1, \dots, n\}$. Hence, the resulting structure of the reduced system is a hypergraph \mathcal{H} with n vertices (Fig. 3c,d and Supplementary Information section IIIC), which is generally directed⁵², weighted, signed and formed from \mathcal{D} , \mathcal{W} and \mathcal{T} .

Beyond the influence of dynamical parameters like the weight matrix W , equation (7) highlights the crucial role of the reduction matrix M in shaping higher-order interactions. Indeed, M partially determines the directed, weighted and signed nature of the hypergraph. Moreover, if the observables depend on disjoint groups of vertices, that is, $M_{\mu i} \propto \delta_{\mu s(i)}$, where δ is the Kronecker delta and s maps each vertex i to its group, then the tensor with elements in equation (7) can be exactly mapped to a matrix. In other words, for the epidemiological dynamics, the higher-order interactions emerge from observables that depend on overlapping groups of vertices (for example, $M = V_n^\top$ in general). Interestingly, such overlapping is a very common characteristic of complex networks such as social networks⁵³.

These observations encouraged us to seek generic conditions for such emergence. For $\dot{x}_i = h_i(x_i, y_i)$, where $h_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ is an analytical scalar field for all $i \in \{1, \dots, N\}$, we proved that the least-square

optimal vector field depends upon higher-order interactions between the observables X_1, \dots, X_n (Methods and Proposition 5). We then deduced two insightful consequences. First, if the scalar field is a polynomial of total degree δ in x_i and y_i for all i , the hypergraph of the reduced system has interactions of maximal order $\delta + 1$ (Methods and Supplementary Corollary 65). Second, having observables depending on disjoint groups of vertices is not sufficient to avoid higher-order interactions in general: the nonlinearity in y_i also plays its part (Methods and Supplementary Corollary 66). Other worked examples for microbial and oscillator dynamics are given in Extended Data Table 1 to complement the previous observations of epidemiological dynamics.

Overall, our results suggest that many instances of higher-order interactions could be a byproduct of the low-dimensional (macroscopic) representation chosen to model a wide variety of complex systems. They clarify the essential role of the description dimension and of the nonlinearity of the original system in shaping the interactions of the ensuing reduced system.

Conclusions and outlook

In this Article, we established the ubiquity of the low-rank hypothesis in complex systems and its consequences, from the dimension reduction of high-dimensional nonlinear dynamics on networks to the emergence of higher-order interactions.

Our experimental results suggest that the low-rank hypothesis is, perhaps, not only a hypothesis but intrinsic to many real complex systems. Our findings hint at the possibility that some emergent collective phenomena are consequences of much fewer variables than what would be expected a priori, thanks to the low-rank nature of their complex network. However, the low-rank hypothesis should be used very carefully: the effective ranks of real networks are often at a non-negligible fraction of N and adopting the low-rank hypothesis unknowingly can lead to an oversimplified model of a given complex system. Thus, it seems relevant to design new random graphs based on the observed singular values of real networks. The singular values of a network are not a mere abstraction from spectral theory: like the degree, the clustering or the reciprocity, they have an intuitive interpretation as indicators of the effective dimension of complex networks and systems.

Our theoretical framework also suggests that inferring the connections in complex systems from time series observed at a relatively coarse-grained resolution (for example, local field potentials in the brain⁵⁴ or abundances in plant communities⁵⁵) would probably reveal significant higher-order interactions. We conjecture that monitoring complex systems at different scales experimentally will clarify the role of the dimension at which the measurements are done for the emergence of higher-order interactions. The dimension reduction of dynamics on higher-order networks^{13,56} is also to be pursued, perhaps through Tucker decomposition⁵⁷.

Nevertheless, determining the precise form of the dominant observables that drive the behaviour of complex systems remains an open problem. Although we have focused on linear observables, there might exist a small set of nonlinear observables well suited for a given high-dimensional dynamics⁵⁸. However, finding appropriate, intuitive, nonlinear observables is much harder⁵⁹. Our observations of the effective ranks of real networks also motivate further research on the inference of interpretable low-rank models from time series⁶⁰.

Finally, one defining property of complex systems that we have not addressed is their capacity for adaptation⁶¹. Our preliminary results suggest that the low effective rank of complex networks plays a central role in controlling^{62,63} and assessing the resilience of complex adaptive systems⁶⁴. This, alongside indications that maturation or learning could reduce a network's effective ranks (Supplementary Information section IIIE and ref. 65), will be the topic of an upcoming publication.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41567-023-02303-0>.

References

- Bellman, R. *Dynamic Programming* (Princeton Univ. Press, 1957).
- Ganguli, S. & Sompolinsky, H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* **35**, 485 (2012).
- Abbott, L. F. et al. The mind of a mouse. *Cell* **182**, 1372 (2020).
- Anderson, P. W. More is different. *Science* **177**, 393 (1972).
- Strogatz, S. et al. Fifty years of ‘More is different’. *Nat. Rev. Phys.* **4**, 508 (2022).
- May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261**, 459 (1976).
- von Neumann, J. The general and logical theory of automata. In *John von Neumann Collected Work*, Vol. V (ed. Taub, A. H.) 288–328 (Pergamon, 1963).
- Wolfram, S. Cellular automata as models of complexity. *Nature* **311**, 419 (1984).
- Parisi, G. Statistical physics and biology. *Phys. World* **6**, 42 (1993).
- Stein, D. L. & Newman, C. M. *Spin Glasses and Complexity* (Princeton Univ. Press, 2013).
- Funahashi, K. I. & Nakamura, Y. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Netw.* **6**, 801 (1993).
- Fortunato, S. & Newman, M. E. J. 20 years of network community detection. *Nat. Phys.* **18**, 848 (2022).
- Bianconi, G. *Higher-Order Networks* (Cambridge Univ. Press, 2021).
- Battiston, F. et al. The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093 (2021).
- Wilf, H. S. The eigenvalues of a graph and its chromatic number. *J. Lond. Math. Soc.* **1**, 330 (1967).
- Donath, W. E. & Hoffman, A. J. Lower bounds for the partitioning of graphs. *IBM J. Res. Dev.* **17**, 420 (1973).
- Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**, 113 (1972).
- Restrepo, J. G., Ott, E. & Hunt, B. R. Onset of synchronization in large networks of coupled oscillators. *Phys. Rev. E* **71**, 036151 (2005).
- Horn, R. A. & Johnson, C. R. *Matrix Analysis* (Cambridge Univ. Press, 2013).
- Weyl, H. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.* **71**, 441 (1912).
- Fan, K. Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Natl Acad. Sci. USA* **37**, 760 (1951).
- Cai, J.-F., Candès, E. J. & Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **46**, 1956 (2010).
- Kutz, J. N., Brunton, S. L. & Brunton, B. W. *Dynamic Mode Decomposition* (SIAM, 2016).
- Gavish, M. & Donoho, D. L. Optimal shrinkage of singular values. *IEEE Trans. Inf. Theory* **63**, 2137 (2017).
- Kalman, R. E. On the general theory of control systems. In *Proc. 1st International IFAC Congress on Automatic and Remote Control* (eds. Coales, J. F., Ragazzini, D. J. & Fuller, A. T.) 491–502 (Elsevier, 1960).
- Kalman, R. E. Contributions to the theory of time-optimal control. *Bol. Soc. Mat. Mex.* **5**, 102 (1960).
- Yan, G. et al. Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* **550**, 519 (2017).
- Marčenko, V. A. & Pastur, L. A. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik* **1**, 457 (1967).
- Féral, D. & Péché, S. The largest eigenvalue of rank one deformation of large Wigner matrices. *Commun. Math. Phys.* **272**, 185 (2007).
- Capitaine, M., Donati-Martin, C. & Féral, D. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuation. *Ann. Probab.* **37**, 1 (2009).
- Benaych-Georges, F. & Nadakuditi, R. R. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227**, 494 (2011).
- Benaych-Georges, F. & Nadakuditi, R. R. The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivar. Anal.* **111**, 120 (2012).
- Pizzo, A., Renfrew, D. & Soshnikov, A. On finite rank deformations of Wigner matrices. *Ann. I. H. Poincaré PR* **49**, 64–94 (2013).
- Baik, J., Ben Arous, G. & Péché, S. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33**, 1643 (2005).
- Valdano, E. & Arenas, A. Exact rank reduction of network models. *Phys. Rev. X* **9**, 031050 (2019).
- Beiran, M., Dubreuil, A., Valente, A., Mastrogiovanni, F. & Ostojević, S. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Comput.* **33**, 1572 (2021).
- Gao, P. & Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32**, 148 (2015).
- Beckermann, B. & Townsend, A. On the singular values of matrices with displacement structure. *SIAM J. Matrix Anal. Appl.* **38**, 1227 (2017).
- Udell, M. & Townsend, A. Why are big data matrices approximately low rank? *SIAM J. Math. Data Sci.* **1**, 144 (2019).
- Gao, J., Barzel, B. & Barabási, A.-L. Universal resilience patterns in complex networks. *Nature* **530**, 307 (2016).
- Tu, C., Grilli, J., Schuessler, F. & Suweis, S. Collapse of resilience patterns in generalized Lotka–Volterra dynamics and beyond. *Phys. Rev. E* **95**, 062307 (2017).
- Jiang, J. et al. Predicting tipping points in mutualistic networks through dimension reduction. *Proc. Natl Acad. Sci. USA* **115**, E639 (2018).
- Laurence, E., Doyon, N., Dubé, L. J. & Desrosiers, P. Spectral dimension reduction of complex dynamical networks. *Phys. Rev. X* **9**, 011042 (2019).
- Végué, M., Thibeault, V., Desrosiers, P. & Allard, A. Dimension reduction of dynamics on modular and heterogeneous directed networks. *PNAS Nexus* **2**, pgad150 (2023).
- Kundu, P., Kori, H. & Masuda, N. Accuracy of a one-dimensional reduction of dynamical systems on networks. *Phys. Rev. E* **105**, 024305 (2022).
- Thibeault, V., St-Onge, G., Dubé, L. J. & Desrosiers, P. Threefold way to the dimension reduction of dynamics on networks: an application to synchronization. *Phys. Rev. Res.* **2**, 043215 (2020).
- Kuehn, C. & Bick, C. A universal route to explosive phenomena. *Sci. Adv.* **7**, 1 (2021).
- St-Onge, G., Thibeault, V., Allard, A., Dubé, L. J. & Hébert-Dufresne, L. Social confinement and mesoscopic localization of epidemics on networks. *Phys. Rev. Lett.* **126**, 098301 (2021).
- Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1 (2020).
- Matheny, M. H. et al. Exotic states in a simple network of nanoelectromechanical oscillators. *Science* **363**, 1057 (2019).
- Nijholt, E., Ocampo-Espindola, J. L., Eroglu, D., Kiss, I. Z. & Pereira, T. Emergent hypernetworks in weakly coupled oscillators. *Nat. Commun.* **13**, 4849 (2022).

52. Gallo, G., Longo, G., Pallottino, S. & Nguyen, S. Directed hypergraphs and applications. *Discret. Appl. Math.* **42**, 177 (1993).
53. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814 (2005).
54. Yu, S. et al. Higher-order interactions characterized in cortical activity. *J. Neurosci.* **31**, 17514 (2011).
55. Mayfield, M. M. & Stouffer, D. B. Higher-order interactions capture unexplained complexity in diverse communities. *Nat. Ecol. Evol.* **1**, 1 (2017).
56. Ferraz de Arruda, G., Tizzani, M. & Moreno, Y. Phase transitions and stability of dynamical processes on hypergraphs. *Commun. Phys.* **4**, 24 (2021).
57. Qi, L. & Luo, Z. *Tensor Analysis* (SIAM, 2017).
58. Watanabe, S. & Strogatz, S. H. Constants of motion for superconducting Josephson arrays. *Phys. D* **74**, 197 (1994).
59. Brunton, S. L., Budišić, M., Kaiser, E. & Kutz, J. N. Modern Koopman theory for dynamical systems. *SIAM Rev.* **64**, 229 (2022).
60. Valente, A., Pillow, J. W. & Ostožić, S. Extracting computational mechanisms from neural data using low-rank RNNs. In *Advances in Neural Information Processing Systems* (eds. Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K. & Oh, A.) 24072–24086 (Curran Associates, Inc., 2022).
61. Holland, J. H. *Hidden Order: How Adaptation Builds Complexity* (Addison-Wesley, 1995).
62. Montanari, A. N., Duan, C., Aguirre, L. A. & Motter, A. E. Functional observability and target state estimation in large-scale networks. *Proc. Natl Acad. Sci. USA* **119**, e2113750119 (2022).
63. Sanhedrai, H. et al. Reviving a failed network through microscopic interventions. *Nat. Phys.* **18**, 338 (2022).
64. Desrosiers, P. & Roy-Pomerleau, X. One for all. *Nat. Phys.* **18**, 238 (2022).
65. Martin, C. H. & Mahoney, M. W. Implicit self-regularization in deep neural networks: evidence from random matrix theory and implications for learning. *J. Mach. Learn. Res.* **22**, 1 (2021).
66. Gower, J. Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra Appl.* **67**, 81 (1985).
67. Gavish, M. & Donoho, D. L. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Trans. Inf. Theory* **60**, 5040 (2014).
68. Donoho, D., Gavish, M. & Johnstone, I. Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Statist.* **46**, 1742 (2018).
69. Scheffer, L. K. et al. A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, 1 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Methods

Random graphs

A random graph can be described by a random matrix

$$W = \langle W \rangle + R, \quad (8)$$

where $\langle W \rangle$ is the expected weight matrix and R is a zero-mean random matrix. Even if one instance in a typical model is generally of full rank N , the expected weight matrix $\langle W \rangle$ is often defined as an element-wise function of a low-rank matrix L , that is,

$$\langle W \rangle = (\phi(L_{ij}))_{i,j=1}^N, \quad (9)$$

where ϕ is a real-valued function of a real variable. This is an alternative, but equivalent, way to write $\langle W \rangle = \phi(L)$ as in the main text. Table 1 lists some classical examples of random graphs and the corresponding low-rank matrices.

In Supplementary Information section IIA, we also report random network models involving two low-rank matrices, such as the general weighted soft configuration model, the general WDSCM and the S^1 weighted RGM, along with other examples (and counterexamples) from network science (for example, the Watts–Strogatz model), random matrix theory, spinglasses, machine learning and neuroscience. Based on these observations and those of ref. 35, one can create many new random graphs with matrices of different ranks.

It is straightforward to assess the low rank of L , but it is harder to assess the low rank of $\langle W \rangle$ when ϕ is nonlinear. For example, in DSCM, $\phi = \phi_{\text{FD}}$, a Fermi–Dirac distribution, and in its weighted version (WDSCM), $\phi = \phi_{\text{BE}}$, a Bose–Einstein distribution. For both models, the following theorem demonstrates that the singular values of their expected weight matrix are bounded above by an exponentially decreasing term.

Theorem 1. This is a simplified version of Supplementary Theorems 32 and 33. Let $\sigma_1 \geq \dots \geq \sigma_N$ be the singular values of $\langle W \rangle$. If $\langle W_{ij} \rangle = \phi_{\text{FD}}(L_{ij}) < 1/2$ or $\langle W_{ij} \rangle = \phi_{\text{BE}}(L_{ij})$ for all $i, j \in \{1, \dots, N\}$, where L is a rank-one matrix, then

$$\sigma_i \leq \sum_{k=i}^{\infty} \ell_k \leq \frac{N\gamma^i}{1-\gamma} \quad \forall i \in \{1, \dots, N\}, \quad (10)$$

where $\ell_k = \sqrt{\sum_{i,j=1}^N L_{ij}^{2k}}$ and $\gamma = \max_{i,j} L_{ij}$.

The proof is based on Weyl's inequalities (Theorem 10 in Supplementary Information section IB) and the truncated geometric series. The bound for $\langle W_{ij} \rangle = \phi_{\text{FD}}(L_{ij}) > 1/2$ is also given in Supplementary Theorem 32. The upper bounds in Theorem 1 expose the low-rank formulation of soft configuration models and paves the way for new bounds on the singular values of other random graphs, such as RGMs.

In Fig. 2, the singular values of W , $\langle W \rangle$ and R are shown for the RPG, DCSBM, S^1 RGM and WDSCM. The upper bounds shown in Fig. 2e, j are given by equation (10), which is computed by summing the constants $n_i > n_{i+1} > \dots$ until n_k is smaller than 10^{-12} . For RPG, the vectors \mathbf{m}_μ and \mathbf{n}_μ are instances of different Gaussian distributions, and $r = 5$. Instances of truncated Pareto distributions were used to generate the expected degrees (DCSBM and S^1 RGM) along with \mathbf{y} and $\bar{\mathbf{y}}$ (WDSCM). The number of blocks q is set to 5 for DCSBM, and the expected number of edges of block matrix L is defined such that there are more edges expected within blocks than between them. To obtain the norm of the random part R of the random weight matrices (except RPG, where R is already set to be a Gaussian of mean 0), we generated 100 instances of W . We computed $R = W - \langle W \rangle$ and then its norm for each instance. The spectral norm of R is increased by changing the variance of each Gaussian element in R for RPG, the expected number of edges in DCSBM, the temperature $1/\beta$ in S^1 RGM, and the minimum values of \mathbf{y} and $\bar{\mathbf{y}}$ in WDSCM. The detailed parameters are given in Supplementary Information section IIA.

Effective ranks

The idea of extracting the number of dominant components in a matrix decomposition is an old theme (for example, in factor analysis^{70,71} or principal component analysis⁷²) but is still subject to new and interesting developments in random matrix theory, data science^{24,67} and in network science, which uses hyperbolic geometry⁷³ and information theory⁷⁴. Because of the close relationship between SVD and the rank, many effective ranks are defined using the singular values. Intuitively, these effective ranks are numbers that indicate how many singular values are important when decomposing a matrix. Table 2 lists the different effective ranks that we have inventoried. The effective ranks srank and shrank are defined from matrix denoising techniques such as those introduced by refs. 24,67,75, which rely on the spectral theory of infinite random matrices³² to determine optimal ways of shrinking the singular values (Supplementary Information section IC). In Fig. 11, the Frobenius norm is used to obtain shrank , and a threshold of 0.9 is used for the energy ratio in Fig. 1j.

As shown in Supplementary Lemma 12, the following ordering of the effective ranks holds: $\text{srank} \leq \text{nrank} \leq \text{erank} \leq \text{rank}$. Because of their simple forms, srank , nrank and erank are amenable to analytic calculations. In particular, we prove that these effective ranks are of order $O(1)$ for singular values with exponentially decreasing envelopes (only stated for srank below).

Corollary 2. This is a simplified version of Supplementary Corollary 40. Let $(W_N)_{N \in \mathbb{Z}_+}$ be an infinite sequence of matrices in which W_N has size $N \times N$. Suppose that there are parameters α and ω such that $0 < \alpha \leq \omega < 1$ and that for each N , the singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0$ of W_N satisfy the inequalities

$$\alpha^{i-1} \leq \frac{\sigma_i}{\sigma_1} \leq \omega^{i-1}, \quad i \in \{1, \dots, N\}. \quad (11)$$

Then, as $N \rightarrow \infty$,

$$\frac{1}{1-\alpha^2} + O(\alpha^{2N}) \leq \text{srank}(W_N) \leq \frac{1}{1-\omega^2} + O(\omega^{2N}). \quad (12)$$

Combined with Theorem 1, the latter corollary implies that the expected weight matrices for the DSCM and its weighted version have $O(1)$ effective ranks.

Moreover, we show in Supplementary Lemma 37 that srank , nrank and erank each have an interpretation in terms of the area under the normalized singular-value scree plots. This point of view allows us to consider a more general family of singular-value envelopes, such as the one in Fig. 1e, to bound the effective ranks. Interestingly, the bounds are related to Gaussian hypergeometric functions, as shown in the next theorem (only stated for srank below, for simplicity).

Theorem 3. This is a simplified version of Supplementary Theorem 38. Suppose that the singular values of matrix W , $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0$, satisfy the inequality

$$\frac{(1-x_i)^{c^*-2}}{(1+\zeta_*x_i)^{b^*}} \leq \frac{\sigma_i}{\sigma_1} \leq \frac{(1-x_i)^{c^*-2}}{(1+\zeta_*x_i)^{b_*}}, \quad (13)$$

where $x_i = (i-1)/(N-1)$ and for some $0 \leq b^- \leq b^+$, $2 \leq c^- \leq c^*$ and $0 < \zeta^- \leq \zeta^+$, and for all $i \in \{1, \dots, N\}$. Then,

$$\frac{N-1}{2c^*-3} H(b^*, c^*, \zeta^*) \leq \text{srank}(W) \leq 1 + \frac{N-1}{2c_*-3} H(b_*, c_*, \zeta_*), \quad (14)$$

where $H(b, c, \zeta) := {}_2F_1(1, 2b; 2(c-1); -\zeta)$ and ${}_2F_1$ is the Gaussian hypergeometric function.

In Fig. 1e, each singular-value distribution of the real networks is interpolated linearly with 1,000 points and the indices are then divided by 1,000. The singular-value envelope is then obtained by fitting the

upper bound in equation (13) to the 95th percentile of the singular values. The fit is done by minimizing the L^2 norm for the parameters $b := b \in [0.01, 10]$, $c := c \in [2, 10]$ and $\zeta := \zeta \in [0.01, 1,000]$ and the minimization gives $b \approx 0.54$, $c \approx 2.3$ and $\zeta \approx 25$. We then use those parameters to evaluate the upper bound in equation (14) divided by N (where we neglect the terms $1/N$), which is shown in Fig. 1f.

Supplementary Corollary 42 shows that if there is a growing graph whose singular values remain bounded within hypergeometric envelopes, then srank , nrank and erank are of order $O(N^{1-\epsilon})$ with $\epsilon \in (0, 1]$ in different asymptotic regimes for the parameters b and ζ , meaning that the effective rank to dimension ratios become negligible asymptotically. Supplementary Information section IIIC clarifies how various singular-value envelopes can lead to very distinct asymptotic behaviours (Supplementary Fig. 4).

When the asymptotic perspective is no longer applicable (for example, for real networks), we cannot classify an effective rank as either low or high. Yet, as explained in the main text, we can use effective rank to dimension ratios, which are well defined for all N , and their values range from 0 (W has rank 0) to 1 (W has full rank).

Dimension reduction of dynamical systems

The dimension reduction of high-dimensional nonlinear dynamics is a fundamental approach for gaining and numerical insights about complex systems. Low-dimensional dynamics can be obtained from an optimization problem in which some error is minimized under a set of constraints to preserve the salient properties of the original system. For dynamical systems, a natural optimization variable is the reduced vector field \mathbf{F} itself, which is chosen to represent approximately the complete vector field \mathbf{f} . Yet, it is rather puzzling to see how the different vector field errors are related to each other and which one can be minimized analytically. In Supplementary Information section IIIB, we provide a useful diagram (Supplementary Diagram 176) that sheds light on different ways to define alignment errors between vector fields.

More precisely, let \mathbf{f} be a complete vector field in \mathbb{R}^N , \mathbf{F} be a reduced vector field in \mathbb{R}^n and M be the $n \times N$ reduction matrix. At $\mathbf{x} \in \mathbb{R}^N$, the alignment error in \mathbb{R}^N is the root-mean-square error (RMSE) between the vector fields \mathbf{f} and $M^+ \circ \mathbf{F} \circ M$,

$$\varepsilon(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - M^+ \mathbf{F}(M\mathbf{x})\|/\sqrt{N}. \quad (15)$$

The alignment error in \mathbb{R}^n is the RMSE between the vector field $M \circ \mathbf{f}$ and $\mathbf{F} \circ M$,

$$\varepsilon(\mathbf{x}) = \|M\mathbf{f}(\mathbf{x}) - \mathbf{F}(M\mathbf{x})\|/\sqrt{n}, \quad (16)$$

where $\|\cdot\|$ is the Euclidean vector norm. When applying the definition of alignment errors on the projected complete vector field $\mathbf{f} \circ P$ instead of \mathbf{f} only, we also define the alignment errors

$$\varepsilon'(\mathbf{x}) = \|\mathbf{f}(P\mathbf{x}) - M^+ \mathbf{F}(M\mathbf{x})\|/\sqrt{N}, \quad (17)$$

$$\varepsilon'(\mathbf{x}) = \|M\mathbf{f}(P\mathbf{x}) - \mathbf{F}(M\mathbf{x})\|/\sqrt{n}, \quad (18)$$

with $P = M^+ M$ being a projector and M^+ being the Moore–Penrose pseudoinverse of M . In principle, the alignment error $\varepsilon(\mathbf{x})$ in \mathbb{R}^N is to be minimized so that it is as close as possible to an exact dimension reduction (Supplementary Definition 47, Supplementary Theorem 48 and Supplementary Diagram 170), but this is far from simple. However, as shown in Supplementary Theorem 52, one can use least squares to show that the vector field of the reduced dynamics

$$\tilde{\mathbf{X}} = M\mathbf{f}(M^+ \mathbf{X}) \quad (19)$$

is optimal in the sense that it minimizes the alignment error $\varepsilon'(\mathbf{x})$ in \mathbb{R}^N . As a consequence, the alignment error $\varepsilon'(\mathbf{x})$ is exactly 0.

In Extended Data Table 1, we show the results of the optimal dimension reduction for five dynamics from different application fields. With the optimal vector field in equation (19) and for dynamics of the general form $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{y})$ (Supplementary Assumptions 70), we find an upper bound on the alignment error $\varepsilon(\mathbf{x})$ related to the singular values of W .

Theorem 4. This is a simplified version of Supplementary Theorem 72. The alignment error $\varepsilon(\mathbf{x})$ in \mathbb{R}^n at $\mathbf{x} \in \mathbb{R}^N$ is upper-bounded as

$$\sqrt{n} \varepsilon(\mathbf{x}) \leq \|V_n^T J'_x(I - V_n V_n^T) \mathbf{x}\| + \sigma_{n+1} \|V_n^T J'_y\|_2 \|\mathbf{x}\|, \quad (20)$$

where $\mathbf{y}' = W\mathbf{x}'$ with \mathbf{x}' being some point between \mathbf{x} and $V_n V_n^T \mathbf{x}$, σ_i is the i th singular value of W , and $J'_x = J_x(\mathbf{x}', \mathbf{y}')$ and $J'_y = J_y(\mathbf{x}', \mathbf{y}')$ are the Jacobian matrices of \mathbf{f} with derivatives according to the vectors \mathbf{x} and \mathbf{y} , respectively. Moreover, for any \mathbf{x} not at the origin of \mathbb{R}^N , the following upper bound holds:

$$\frac{\varepsilon(\mathbf{x})}{\|\mathbf{x}\|} \leq \frac{1}{\sqrt{n}} [\alpha(\mathbf{x}', \mathbf{y}') + \sigma_{n+1} \beta(\mathbf{x}', \mathbf{y}')], \quad (21)$$

where $\alpha(\mathbf{x}', \mathbf{y}') = \sigma_1(J_x(\mathbf{x}', \mathbf{y}'))$ and $\beta(\mathbf{x}', \mathbf{y}') = \sigma_1(J_y(\mathbf{x}', \mathbf{y}'))$.

As a bonus, the proof of the theorem suggests choosing M as the truncated right singular vectors V_n^T , since it allows a part of the bound to be minimized. This is a consequence of the Schmidt–Eckart–Young–Mirsky theorem and more specifically, Supplementary Theorem 9. This choice for M also has a notable consequence: each observable X_μ generally becomes a global observable in that it contains information about most vertices. This characteristic, alongside that it is a finite-size dimension reduction, makes our approach stand out from many mean-field modelling approaches used in network science in which vertices are coarse-grained according to their degree (local property) or to some other mesoscopic property of the network.

Theorem 4 also provides a criterion for exact dimension reduction: if $J_x(\mathbf{x}', \mathbf{y}') = dI$ for some real constant d and n is the rank of W , then $\varepsilon(\mathbf{x}) = 0$ (Supplementary Corollary 74 in Supplementary Information section IIID). For example, consider the class of dynamics in matrix form

$$\dot{\mathbf{x}} = d\mathbf{x} + \mathbf{s}(W\mathbf{x}), \quad (22)$$

where \mathbf{s} is a vector of N functions $s_i : \mathbb{R} \rightarrow \mathbb{R}$ and W has rank r and compact SVD $U_r \Sigma_r V_r^T$. This can be exactly reduced to the r -dimensional reduced dynamics:

$$\dot{\tilde{\mathbf{X}}} = d\tilde{\mathbf{X}} + V_r^T \mathbf{s}(U_r \Sigma_r \mathbf{X}), \quad (23)$$

where $\mathbf{X} = V_r^T \mathbf{x}$. For any n and $\mathbf{X} = V_n^T \mathbf{x}$, the vector field in equation (23) is the least-squares optimal one in the sense described in Theorem 52 in Supplementary Information section IIIB. This result implies that any RNN or any neuronal dynamics (with $a = 0$) having the forms given in Extended Data Table 1 can be exactly reduced (examples 76 and 77 in Supplementary Information section IIID).

A simple corollary of Theorem 4 (Supplementary Corollary 79) shows that if the dynamics is that of a linear system, the relative alignment error in \mathbb{R}^n at $\mathbf{x} \in \mathbb{R}^N$ is

$$\frac{\varepsilon(\mathbf{x})}{\|\mathbf{x}\|} \leq \frac{\sigma_{n+1}}{\sqrt{n}}, \quad (24)$$

implying that a rapid decrease of the singular values of W directly induces a rapid decrease of the alignment error.

Emergence of higher-order interactions

All the N -dimensional (complete) dynamics on a network in Extended Data Table 1 (and many more; see Supplementary Information section

III C) have the general form $\dot{x}_i = h_i(x_i, y_i)$ for all $i \in \{1, \dots, N\}$, where $x_i : [0, \infty) \rightarrow \mathbb{R}$, $y_i = \sum_{j=1}^N W_{ij}x_j$ and $h_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ is an analytic function.

Proposition 5. This is a simplified version of Supplementary Proposition 61. The least-square reduced dynamics can be expressed in terms of higher-order interactions between the observables as

$$\begin{aligned}\dot{X}_{\mu} &= c_{\mu} + \sum_{d_x=1}^{\infty} \sum_{\alpha} \mathcal{D}_{\mu\alpha}^{(d_x+1)} X_{\alpha} + \sum_{d_y=1}^{\infty} \sum_{\beta} \mathcal{W}_{\mu\beta}^{(d_y+1)} X_{\beta} \\ &+ \sum_{d_x, d_y=1}^{\infty} \sum_{\alpha, \beta} \mathcal{T}_{\mu\alpha\beta}^{(d_x+d_y+1)} X_{\alpha\beta},\end{aligned}$$

where we have introduced the multi-indices $\alpha = (\alpha_1, \dots, \alpha_{d_x})$ and $\beta = (\beta_1, \dots, \beta_{d_y})$ with $\alpha_p, \beta_q \in \{1, \dots, n\}$, the compact notation for products $X_{\gamma} = X_{\gamma_1} \dots X_{\gamma_d}$ whereas c_{μ} denotes a real constant and $\mu \in \{1, \dots, n\}$. The higher-order interactions are described by three tensors of respective order $d_x + 1, d_y + 1$ and $d_x + d_y + 1$ whose elements are

$$\begin{aligned}\mathcal{D}_{\mu\alpha}^{(d_x+1)} &= \sum_{i=1}^N c_{id_x, 0} M_{\mu i} M_{i\alpha}^+, \\ \mathcal{W}_{\mu\beta}^{(d_y+1)} &= \sum_{i=1}^N \sum_{j} c_{i0d_y} M_{\mu i} W_{ij} M_{j\beta}^+, \\ \mathcal{T}_{\mu\alpha\beta}^{(d_x+d_y+1)} &= \sum_{i=1}^N \sum_{j} c_{id_x, d_y} M_{\mu i} M_{i\alpha}^+ W_{ij} M_{j\beta}^+,\end{aligned}$$

for some real coefficients c_{id_x, d_y} with $i \in \{1, \dots, N\}$, $d_x, d_y \in \mathbb{Z}_+$ and the multi-index j in the sums is in $\{1, \dots, N\}^{d_y}$.

This proposition led us to two corollaries. First, if $h_i(x_i, y_i)$ is a polynomial of total degree δ in x_i and y_i , then the reduced dynamics has a polynomial vector field of total degree δ with interactions of maximal order $\delta + 1$ (Supplementary Corollary 65). Second, if M is block diagonal and h_i linearly depends on y_i , then there are solely pairwise interactions in the reduced system, which does not hold in general for nonlinear dependencies of h_i over y_i (Supplementary Corollary 66).

In Extended Data Table 1, we apply Proposition 5 to the quenched mean-field susceptible-infected-susceptible dynamics, the microbial dynamics and the Kuramoto–Sakaguchi dynamics. The results concretely illustrate the emergence of higher-order interactions through dimension reduction. More details are given in Supplementary Information section IIIC.

Integration and properties of the dynamics

The trajectories of the dynamics on the real networks presented in Fig. 4 were obtained with `solve_ivp` from `scipy.integrate`. We used the backward differentiation formula, an implicit method with variable step length and order, which is known to be well suited for stiff problems, such as the microbial dynamics on the gut microbiome. We observed that a relative tolerance $rtol = 10^{-8}$ and an absolute tolerance $atol = 10^{-12}$ for the complete microbial dynamics ($rto = 10^{-6}$ and $atol = 10^{-10}$ for the reduced dynamics) gave reliable results with decent integration time while being in line with the recent benchmarks of ref. 76. Moreover, we have provided the Jacobian matrices of the complete and reduced dynamics to the integrator as recommended in the documentation of `solve_ivp` for the backward differentiation formula. We also integrated the other dynamics with backward differentiation with a relative tolerance of 10^{-8} and an absolute tolerance of 10^{-12} .

For the epidemiological dynamics, the critical slowing down appears but is easily dealt with by increasing the number of time steps near the transcritical bifurcation (at the infection rate of 1, that is, the largest singular value of the rescaled network), as we have done in the inset of Fig. 4e. Note that increasing the dimension improves the prediction for higher infection rates. In Fig. 4f, we observe a hysteresis for the global observable of the neuronal dynamics versus the synaptic weight. In Fig. 4e,f, the RMSEs are simply computed between the

global equilibrium points of the complete and the reduced dynamics at different n .

As illustrated in Fig. 4g, several branches of stable equilibrium points for the global observables of the microbial dynamics arise. We proceeded as follows to get a simplified picture involving only some equilibrium point branches. We focused on one forward branch obtained with initial conditions \mathbf{x}_0 sampled from a uniform distribution between 0 and 1, and in Fig. 4g, we showed its loss of stability when incrementally increasing the microbial interaction weight. To obtain one backward branch, we sampled the initial condition \mathbf{x}_0 from a uniform distribution between 0 and z , where z is a random integer between 1 and 15. We integrated the dynamics to get the equilibrium point. Next, we decreased the microbial interaction weight and used the last equilibrium point as the initial condition for the integration and repeated these last two steps until the minimum coupling value (0.1 in Fig. 4g) was reached. We repeated all these steps 100 times (300 for $n = 76$) to generate different initial conditions and stable branches. At each iteration, we ensured that the vector fields evaluated at the equilibrium points gave a vector with elements below the tolerance 10^{-7} and that the equilibrium points were positive (Supplementary Information section IIIH). In this case, the RMSE was computed between the average upper and lower branches of the complete and reduced dynamics.

For a (finite-sized) RNN, like the observations in the conclusion of ref. 77, there is a stable equilibrium point at zero for lower coupling. Increasing the coupling eventually gives rise to limit cycles of increasing complexity, such as the one in Fig. 4h. We illustrate a three-dimensional projection of this high-dimensional limit cycle in the complete dynamics and those in the reduced dynamics as the dimension n approaches the rank of the learned network. The RMSE was computed between the points of the limit cycle for the complete recurrent neural dynamics and the closest points on the limit cycles of the reduced dynamics.

The choices of global observables used in Fig. 4 are justified in Supplementary Information section IIIF and the parameter values of the dynamics are in Extended Data Table 1.

Data availability

All the details about the real networks data used in the paper, mostly from the network repository Netzschleuder, are given in Supplementary Information section IV. The data to generate Figs. 1, 2 and 4 are available on Zenodo (<https://doi.org/10.5281/zenodo.8342130>).

Code availability

The Python code used to generate the results of the paper is available on Zenodo (<https://doi.org/10.5281/zenodo.8342130>). The code for the optimal shrinkage of singular values is a Python implementation of the Matlab codes `optimal_singval_threshold` (ref. 67) and `optimal_singval_shrink` (ref. 24), which is partly based on the repository `optpt` by B. Erichson.

References

70. Malinowski, E. R. Theory of error in factor analysis. *Anal. Chem.* **49**, 606 (1977).
71. Sánchez, E. & Kowalski, B. R. Generalized rank annihilation factor analysis. *Anal. Chem.* **58**, 496 (1986).
72. Abdi, H. & Williams, L. J. Principal component analysis. *WIREs Comput. Stat.* **2**, 433 (2010).
73. Almagro, P., Boguñá, M. & Ángeles Serrano, M. Detecting the ultra low dimensionality of real networks. *Nat. Commun.* **13**, 6096 (2022).
74. Lynn, C. W. & Bassett, D. S. Compressibility of complex networks. *Proc. Natl Acad. Sci. USA* **118**, e2023473118 (2021).
75. Perry, P. O. *Cross-Validation for Unsupervised Learning*. PhD thesis, Stanford Univ. (2009).

76. Städter, P., Schälte, Y., Schmiester, L., Hasenauer, J. & Stapor, P. L. Benchmarking of numerical integration methods for ODE models of biological systems. *Sci. Rep.* **11**, 2696 (2021).
77. Sompolinsky, H., Crisanti, A. & Sommers, H.-J. Chaos in random neural networks. *Phys. Rev. Lett.* **61**, 259 (1988).

Acknowledgements

We are grateful to G. Eilerstein for sharing the code to extract the weight matrices from the repository NWS, G. Jékely for sharing the neuronal and desmosomal connectomes of *Platynereis dumerilii*, C. Murphy for useful discussions on artificial neural networks, G. St-Onge for his comments on the preprint and X. Roy-Pomerleau for helping us to explore the microbial dynamics numerically. We thank É. Boran for his fundamental contribution to linear algebra. This work was supported by the Fonds de recherche du Québec—Nature et technologies (V.T. and P.D.), the Natural Sciences and Engineering Research Council of Canada (V.T., A.A. and P.D.) and the Sentinelle Nord programme of Université Laval, funded by the Canada First Research Excellence Fund (V.T., A.A. and P.D.).

Author contributions

All authors contributed to the formulation of the study, the interpretation of the results and the editing of the paper. V.T. and P.D. obtained the mathematical results and conceived the conceptual basis of the project. V.T. led the writing of the manuscript, wrote the

supplementary information with P.D., designed the figures, wrote the code and performed the numerical experiments to generate the results. V.T., A.A. and P.D. contributed to the code and analysed the data to generate Fig. 1.

Competing interests

The authors declare no competing interests.

Additional information

Extended data Extended data are available for this paper at <https://doi.org/10.1038/s41567-023-02303-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41567-023-02303-0>.

Correspondence and requests for materials should be addressed to Vincent Thibeault or Patrick Desrosiers.

Peer review information *Nature Physics* thanks Jianxi Gao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Extended Data Table 1 | Vector fields of typical nonlinear dynamics on a network and their least-square optimal reduced dynamics

| Name | Complete vector field $h_i(x_i, y_i)$ | Reduced vector field $H_\mu(X_1, \dots, X_n)$ |
|-----------------|--|---|
| Epidemiological | $-d_i x_i + \gamma(1 - x_i) y_i$ | $\sum_{\nu=1}^n (\mathcal{D}_{\mu\nu} + \mathcal{W}_{\mu\nu}) X_\nu + \sum_{\nu, \kappa=1}^n \mathcal{T}_{\mu\nu\kappa} X_\nu X_\kappa$ |
| Microbial | $a - d_i x_i + b x_i^2 - c x_i^3 + \gamma x_i y_i$ | $\mathcal{C}_\mu + \sum_{\nu=1}^n \mathcal{D}_{\mu\nu} X_\nu + \sum_{\nu, \kappa=1}^n (\mathcal{D}_{\mu(\nu, \kappa)} + \mathcal{T}_{\mu\nu\kappa}) X_\nu X_\kappa + \sum_{\nu, \kappa, \tau=1}^n \mathcal{D}_{\mu(\nu, \kappa, \tau)} X_\nu X_\kappa X_\tau$ |
| Oscillator | $i\omega_i x_i + \gamma e^{-i\alpha} y_i - \gamma e^{i\alpha} x_i^2 \bar{y}_i$ | $\sum_{\nu=1}^n (\mathcal{D}_{\mu\nu} + \mathcal{W}_{\mu\nu}) X_\nu + \sum_{\nu, \kappa, \tau=1}^n \mathcal{T}_{\mu(\nu, \kappa, \tau)} X_\nu X_\kappa \bar{X}_\tau$ |
| RNN | $-d_i x_i + \tanh(\gamma y_i + c_i)$ | $\sum_{\nu=1}^n \mathcal{D}_{\mu\nu} X_\nu + \sum_{i=1}^N M_{\mu i} \tanh\left(\gamma \sum_{\nu=1}^n \mathcal{W}_{j\nu} X_\nu + c_i\right)$ |
| Neuronal | $-d_i x_i + (1 - a x_i) \mathcal{S}[b(\gamma y_i - c_i)]$ | $\sum_{\nu=1}^n \mathcal{D}_{\mu\nu} X_\nu + \sum_{j=1}^N M_{\mu j} \left(1 - a \sum_{\nu=1}^n M_{j\nu}^+ X_\nu\right) \mathcal{S}\left[b\left(\gamma \sum_{\kappa=1}^n \mathcal{W}_{j\kappa} X_\kappa - c_i\right)\right]$ |

The epidemiological dynamics is the quenched mean-field susceptible-infected-susceptible (QMF SIS). Its parameters in Fig. 4e are $d_i=1$ for all i , y is the global infection rate and the global observable $\mathcal{X}^ = \mathbf{w} \cdot \mathbf{X}^*$ is defined with $\mathbf{w} = (w_1 \ 0 \dots \ 0)^\top$ with $w_1 = 1/\sum_{j=1}^N (\mathbf{v}_1)_j$, where \mathbf{v}_1 is the leading right singular vector of the contact network. The microbial dynamics is a population dynamics and when $a=b=c=0$, it is the well-known generalized Lotka-Volterra dynamics. Its parameters in Fig. 4g are $a=5$, $b=13$, $c=1$ and $d_i=30$ for all i , y is the global microbial interaction weight, and the global observable is defined with $\mathbf{w} = (w_1 \ \dots \ w_{n_{\min}} \ 0 \ \dots \ 0)^\top$ where $\tilde{\mathbf{w}} = (w_1 \ \dots \ w_{n_{\min}}) = V_{n_{\min}}^\top \mathbf{1}/(10N)$ and $n_{\min} = 76$, which is chosen to approach the uniform global observable $\sum_{i=1}^N x_i/(10N)$. The oscillator dynamics is the Kuramoto-Sakaguchi dynamics for which $x_i = e^{i\theta_i}$, where θ_i is the phase of the oscillator, i is the imaginary unit (to be distinguished from the index i), an underscore denotes complex conjugation, and X_1, \dots, X_n are, thus, complex. The parameters of the RNN in Fig. 4h are $d_i=1$ for all i and $y=2$. The neuronal dynamics is the Wilson-Cowan dynamics, where \mathcal{S} denotes the sigmoid function $x \mapsto 1/(1+e^{-x})$. Its parameters in Fig. 4f are $a=0.05$, $b=1$, $c_i=1$, $d_i=1$ for all i , y is the global synaptic weight and the global observable is defined with $\mathbf{w} = (w_1 \ 0 \ \dots \ 0)^\top$ with $w_1 = 1/(0.15 \sum_{j=1}^N (\mathbf{v}_1)_j)$ where \mathbf{v}_1 is the leading right singular vector of the connectome. More details about the dynamics and their parameters are given in Supplementary Sections IIIB and IIIC.