nature physics

Article

https://doi.org/10.1038/s41567-023-02303-0

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 equation[6], 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 systems[11].

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 modules[12]. 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 invariants[15], their modular structure[16], the centrality of their vertices[17] or the bifurcations of a dynamical system represented by a network[18].

[1]Département de physique, de génie physique et d'optique, Université Laval, Québec, Québec, Canada. [2]Centre interdisciplinaire en modélisation mathématique de l'Université Laval, Québec, Québec, Canada. [3]Centre de recherche CERVO, Québec, Québec, Canada.
e-mail: vincent.thibeault.1@ulaval.ca; patrick.desrosiers@phy.ulaval.ca

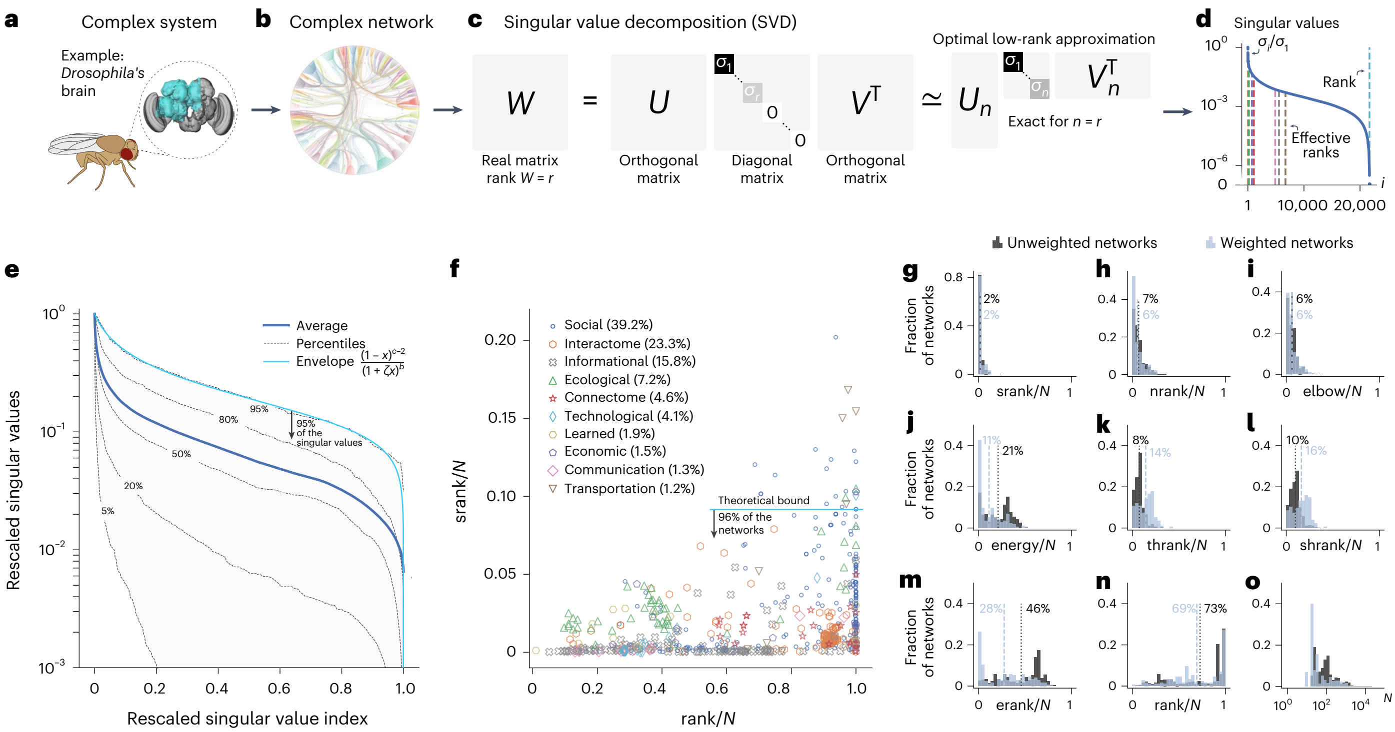

Fig. 1 | Experimental verification of the low-rank hypothesis for real networks. **a**, *Drosophila melanogaster*'s hemibrain is an example of a complex system. The open-source image of the hemibrain is from ref. 69. **b**, A complex network illustration of *Drosophila melanogaster*'s connectome[69]. 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 nrank (**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 $\sigma_1, \ldots, \sigma_N$ 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 decomposition[19], 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 completion[22], dynamic mode decomposition[23] and optimal singular-value shrinkage[24]), control theory (for example, the Kalman criterion[25–27]), random matrix theory (for example, Marčenko–Pastur's law[28]) and linear algebra (for example, matrix norms[19]). 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,

https://doi.org/10.1038/s41567-023-02303-0

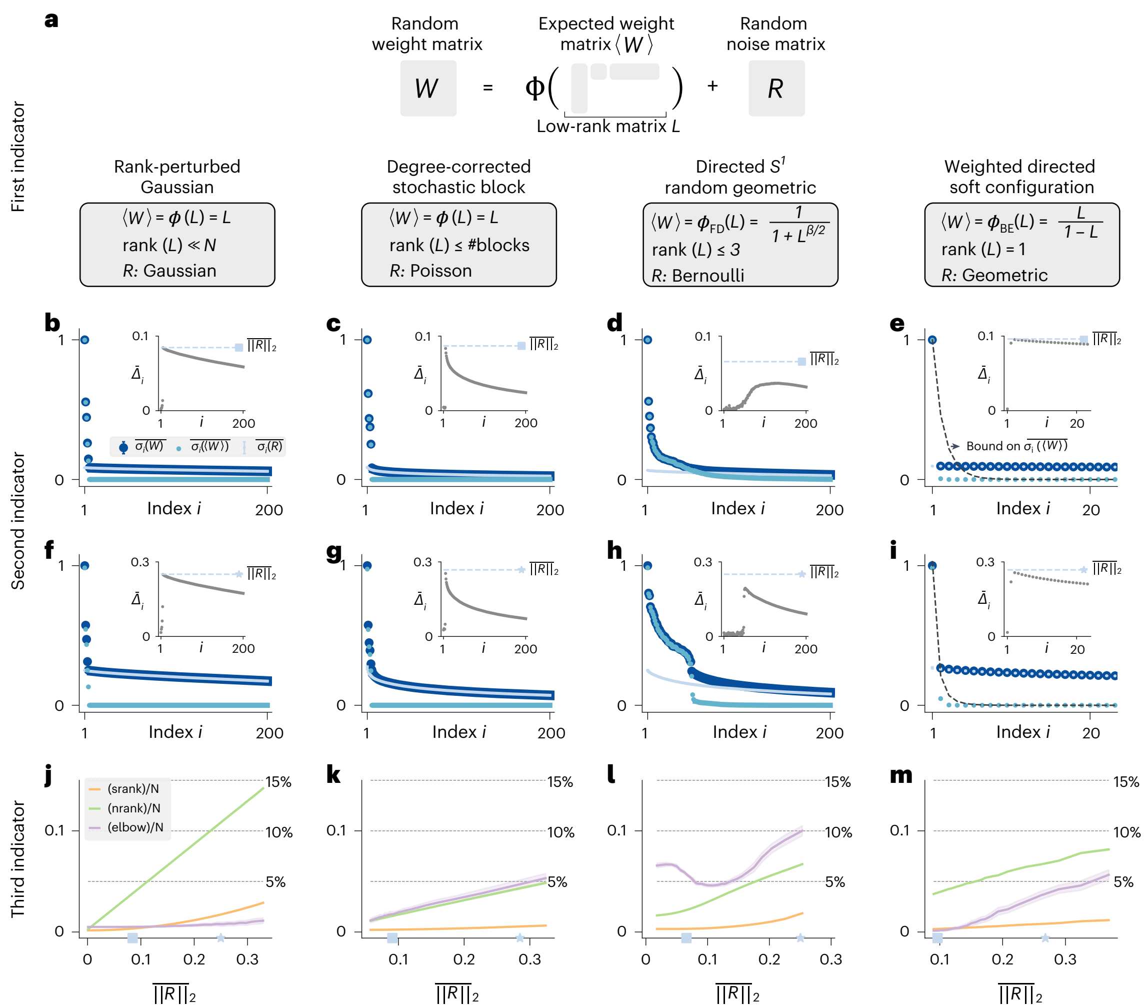

Fig. 2 | Three indicators of the low-rank hypothesis for random graphs. **a**, Many random graphs have a random matrix representation in which the expected weight matrix $\langle W\rangle$ is a matrix-valued function Φ of a low-rank matrix L plus a centred random part R. Four examples of random matrices 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 $\Phi_{\rm FD}$ and $\Phi_{\rm BE}$, 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 $\overline{\|R\|_2}$ near 0.1 (**b**–**e**) and star markers for $\overline{\|R\|_2}$ near 0.3 (**f**–**i**). The singular values are respectively denoted $\overline{\sigma_i(W)}$, $\overline{\sigma_i(\langle W\rangle)}$ and $\overline{\sigma_i(R)}$ (from darker to lighter blue markers) where $\overline{x} = \langle x\rangle/\langle \|W\|_2\rangle$ and $\langle\,\rangle$ 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 $i \in \{1, \ldots, N\}$ 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,

$$\Delta_i = |\sigma_i(W) - \sigma_i(\langle W\rangle)| \le \|R\|_2, \tag{1}$$

for all $i \in \{1, \ldots, N\}$, where $\sigma_i(A)$ denotes the ith 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 $\mathrm{rank}(L) = r$ 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 transition[34]. 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 $\sigma_1(\langle W\rangle), \ldots, \sigma_r(\langle W\rangle)$ (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	$\mathcal{G}(N,p)$	$Np\,\hat{\mathbf{1}}\hat{\mathbf{1}}^\top$	1	L_{ij}
	CL	$\frac{\lVert\kappa\rVert^2}{2M}\hat{\boldsymbol{\kappa}}\hat{\boldsymbol{\kappa}}^\top$	1	L_{ij}
	DSCM	$\lVert\boldsymbol{\alpha}\rVert\,\lVert\boldsymbol{\beta}\rVert\,\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}^\top$	1	$\frac{L_{ij}}{1+L_{ij}}$
	MD	$\sum_{\mu,\nu=1}^{r}\Delta_{\mu\nu}\,\mathbf{v}_\mu\mathbf{v}_\nu^\top$	r	L_{ij}
	SBM	$\sum_{\mu,\nu=1}^{q}\sqrt{n_\mu n_\nu}\,p_{\mu\nu}\,\mathbf{b}_\mu\mathbf{b}_\nu^\top$	$\leq q$	L_{ij}
	S^D RGM	$\frac{R^2}{\mu^2}\left(\bar{\boldsymbol{\kappa}}_{\text{in}}\bar{\boldsymbol{\kappa}}_{\text{out}}^\top\right)\circ\bar{\theta}$	$\leq D+2$	$\frac{1}{1+L_{ij}^{\beta/2}}$
Weighted	$\mathcal{G}(N,p,w)$	$Npw\,\hat{\mathbf{1}}\hat{\mathbf{1}}^\top$	1	L_{ij}
	WCL	$\mathbf{y}\mathbf{y}^\top$	1	L_{ij}
	WDSCM	$\mathbf{y}\bar{\mathbf{y}}^\top$	1	$\frac{L_{ij}}{1-L_{ij}}$
	RPG	$\sum_{\mu=1}^{r}\mathbf{m}_\mu\mathbf{n}_\mu^\top$	r	L_{ij}
	WSBM	$\sum_{\kappa,\nu=1}^{q}\sqrt{n_\kappa n_\nu}\,\mu_{\kappa\nu}\,\mathbf{b}_\kappa\mathbf{b}_\nu^\top$	$\leq q$	L_{ij}
	DCSBM	$\Lambda\circ\left(\hat{\boldsymbol{\kappa}}_{\text{in}}\hat{\boldsymbol{\kappa}}_{\text{out}}^\top\right)$	$\leq q$	L_{ij}
	RDPG	$\sum_{\mu=1}^{d}\mathbf{X}_\mu\mathbf{X}_\mu^\top$	$\leq d$	L_{ij}

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 IIA.

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

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

For energy, the constant τ is a threshold to be set between 0 and 1. For thrank, σ_{med} is the median singular value and μ_{med} is the median of a Marčenko–Pastur probability density function[67]. 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.

are in the vicinity of $\sigma_1(\langle W\rangle),\ldots,\sigma_r(\langle W\rangle)$, 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 $N\to\infty$ (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 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 srank ≤ nrank ≤ erank ≤ rank

 https://doi.org/10.1038/s41567-023-02303-0

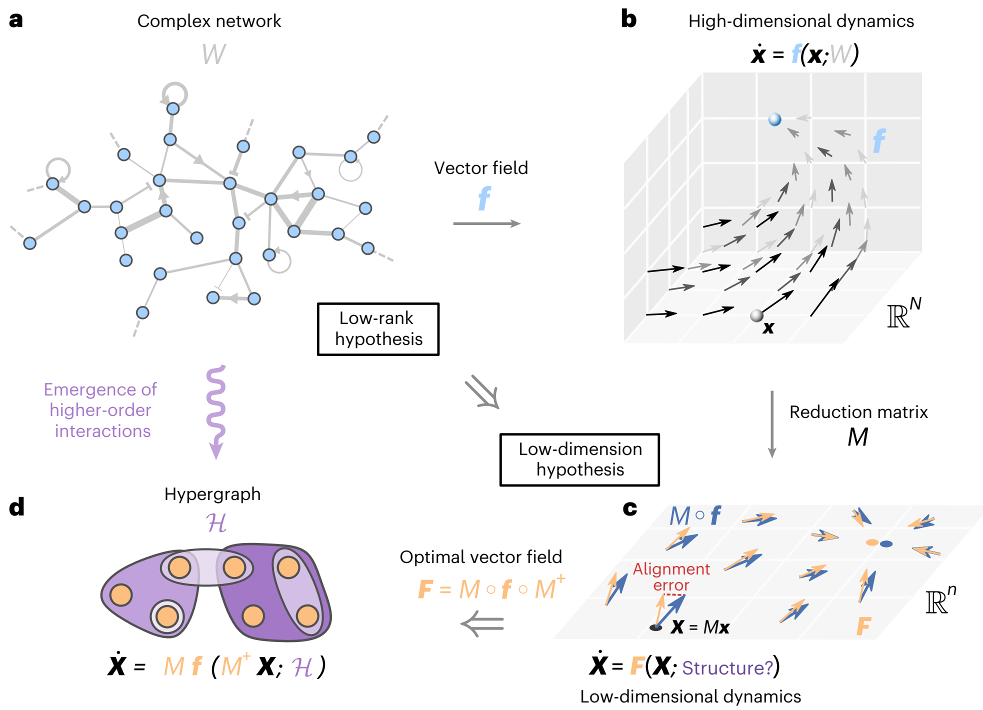

Fig. 3 | The low-rank hypothesis of complex systems and the emergence of higher-order interactions. a, A complex network represented as a weighted (edge widths), signed and directed (edges with arrows or a perpendicular line for inhibition) graph with weight matrix W. **b**, A vector field **f** of an 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 $M : \mathbb{R}^N \to \mathbb{R}^n$, 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 **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 $X_1, ..., X_n$ 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 IIIC) 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. 1n). 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. 1e 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, $\mathbf{f} : \mathbb{R}^N \to \mathbb{R}^N$ is a continuously differentiable vector field and W is an $N \times N$ weight matrix describing the network (Fig. 3a,b). More specifically, given $\mathbf{g} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ 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 IIIA). 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 **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 IIIA), 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 observables[46]), might further complicate the optimization problem.

Let us concentrate on identifying **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

https://doi.org/10.1038/s41567-023-02303-0

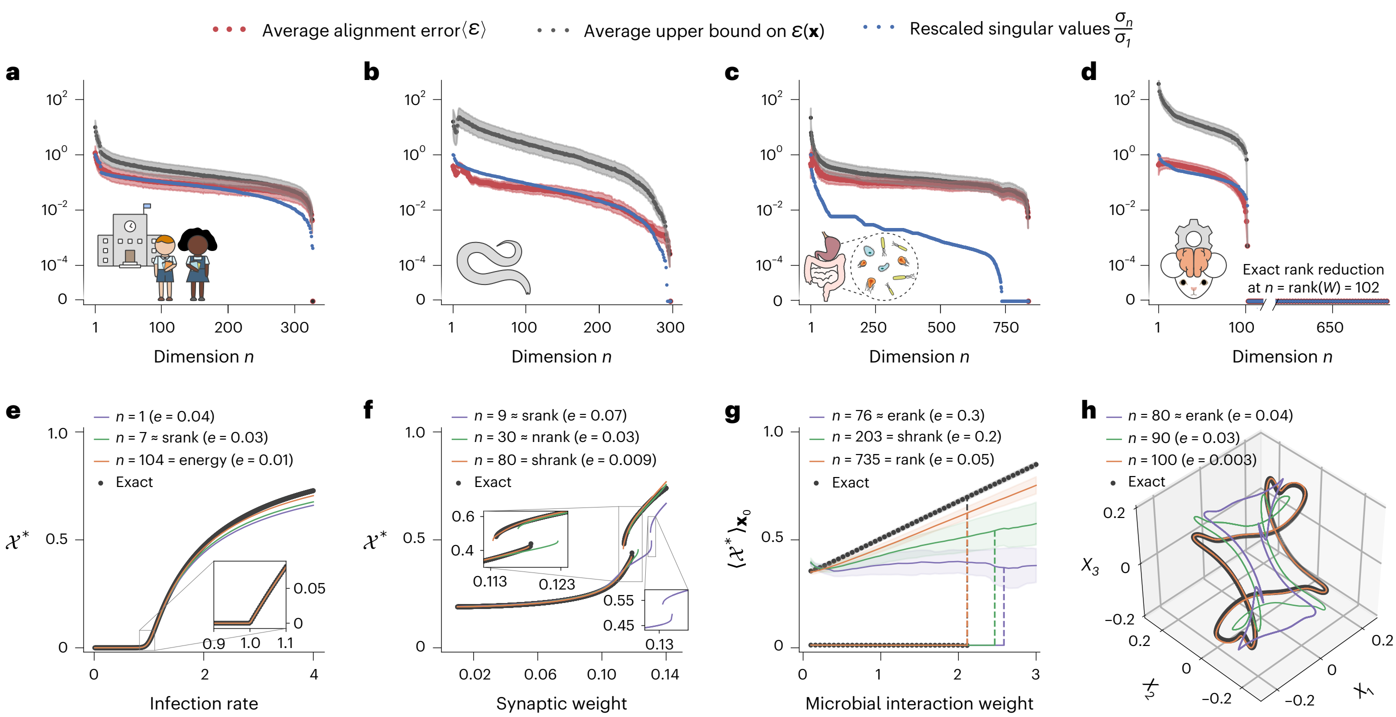

Fig. 4 | Dimension reduction errors for nonlinear dynamics on real complex networks related to their singular values and effective ranks. a–d, The decrease of the alignment error $\mathcal{E}(\mathbf{x})$ (red markers) is in accordance with the rapid decrease of singular values (blue markers), as expected 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 $\mathbf{x}$ 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 $\mathcal{X}^* = \mathbf{w} \cdot \mathbf{X}^*$ at equilibrium where $\mathbf{w}$ 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 $\boldsymbol{x}_0$ (Methods). Here we show the upper branches averaged over $\mathbf{x}_0$ (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 norm[24] 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 $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{y})$, that the alignment error $\mathcal{E}(\mathbf{x})$ caused by the least-square vector field satisfies

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

where $J'_{\mathbf{x}}$ and $J'_{\mathbf{y}}$ 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):

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

where $P = V_n V_n^{\top}$. Notably, the inequality provides a criterion for exact dimension reduction: if $J'_{\mathbf{x}} = dI$ for $d \in \mathbb{R}$ and $n = \mathrm{rank}(W)$, 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 $\mathcal{E}(\mathbf{x})/\|\mathbf{x}\|$ for the linear system $\dot{\mathbf{x}} = W\mathbf{x}$ is simply upper-bounded by $\sigma_{n+1}/\sqrt{n}$, 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 transitions[47] or mesoscopic localization[48]—have been extensively studied[49]. 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 $i \in \{1, ..., N\}$, where x_i is the probability that the vertex i is infected, $\mathbf{y} = W\mathbf{x}$ whereas d_i and γ denote the recovery rate of vertex i and the infection rate, respectively. The reduced dynamics is then given by

$$\dot{X}_\mu = \sum_{\nu=1}^{n} (\mathcal{D}_{\mu\nu} + \mathcal{W}_{\mu\nu}) X_\nu - \gamma \sum_{i=1}^{N} M_{\mu i} \left(\sum_{\nu=1}^{n} M^{+}_{i\nu} X_\nu \right) \left(\sum_{j=1}^{N} \sum_{\kappa=1}^{n} W_{ij} M^{+}_{j\kappa} X_\kappa \right) \quad (5)$$

for all $\mu \in \{1, ..., n\}$, where $\mathcal{D} = -MDM^{+}$ is a reduced $n \times n$ recovery rate matrix with $D = \mathrm{diag}(d_1, ..., d_N)$, 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 μth observable, $M^{\top}_{i\nu} X_\nu$ is the influence of the νth observable weighted by its dependence over vertex i and $W_{ij} M^{\top}_{j\kappa} X_\kappa$ is the influence of the κth 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_ν and X_κ, an observation that is made more explicit by rearranging equation (5) as

$$\dot{X}_\mu = \sum_{\nu=1}^{n} (\mathcal{D}_{\mu\nu} + \mathcal{W}_{\mu\nu}) X_\nu + \sum_{\nu,\kappa=1}^{n} \mathcal{T}_{\mu\nu\kappa} X_\nu X_\kappa, \quad (6)$$

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

$$\mathcal{T}_{\mu\nu\kappa} = -\gamma \sum_{i,j=1}^{N} M_{\mu i} M^{+}_{i\nu} W_{ij} M^{+}_{j\kappa} \quad (7)$$

for $\mu, \nu, \kappa \in \{1, ..., n\}$. 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 directed[52], 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 networks[53].

These observations encouraged us to seek generic conditions for such emergence. For $\dot{x}_i = h_i(x_i, y_i)$, where $h_i : \mathbb{R}^2 \to \mathbb{R}$ is an analytical scalar field for all $i \in \{1, ..., N\}$, we proved that the least-square optimal vector field depends upon higher-order interactions between the observables $X_1, ..., X_n$ (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 brain[54] or abundances in plant communities[55]) 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 decomposition[57].

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 dynamics[58]. However, finding appropriate, intuitive, nonlinear observables is much harder[59]. Our observations of the effective ranks of real networks also motivate further research on the inference of interpretable low-rank models from time series[60].

Finally, one defining property of complex systems that we have not addressed is their capacity for adaptation[61]. 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 systems[64]. This, alongside indications that maturation or learning could reduce a network's effective ranks (Supplementary Information section IIE 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

1. Bellman, R. *Dynamic Programming* (Princeton Univ. Press, 1957).
2. Ganguli, S. & Sompolinsky, H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* **35**, 485 (2012).
3. Abbott, L. F. et al. The mind of a mouse. *Cell* **182**, 1372 (2020).
4. Anderson, P. W. More is different. *Science* **177**, 393 (1972).
5. Strogatz, S. et al. Fifty years of 'More is different'. *Nat. Rev. Phys.* **4**, 508 (2022).
6. May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261**, 459 (1976).
7. 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).
8. Wolfram, S. Cellular automata as models of complexity. *Nature* **311**, 419 (1984).
9. Parisi, G. Statistical physics and biology. *Phys. World* **6**, 42 (1993).
10. Stein, D. L. & Newman, C. M. *Spin Glasses and Complexity* (Princeton Univ. Press, 2013).
11. Funahashi, K. I. & Nakamura, Y. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Netw.* **6**, 801 (1993).
12. Fortunato, S. & Newman, M. E. J. 20 years of network community detection. *Nat. Phys.* **18**, 848 (2022).
13. Bianconi, G. *Higher-Order Networks* (Cambridge Univ. Press, 2021).
14. Battiston, F. et al. The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093 (2021).
15. Wilf, H. S. The eigenvalues of a graph and its chromatic number. *J. Lond. Math. Soc.* **1**, 330 (1967).
16. Donath, W. E. & Hoffman, A. J. Lower bounds for the partitioning of graphs. *IBM J. Res. Dev.* **17**, 420 (1973).
17. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**, 113 (1972).
18. Restrepo, J. G., Ott, E. & Hunt, B. R. Onset of synchronization in large networks of coupled oscillators. *Phys. Rev. E* **71**, 036151 (2005).
19. Horn, R. A. & Johnson, C. R. *Matrix Analysis* (Cambridge Univ. Press, 2013).
20. Weyl, H. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.* **71**, 441 (1912).
21. Fan, K. Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Natl Acad. Sci. USA* **37**, 760 (1951).
22. Cai, J.-F., Candès, E. J. & Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **46**, 1956 (2010).
23. Kutz, J. N., Brunton, S. L. & Brunton, B. W. *Dynamic Mode Decomposition* (SIAM, 2016).
24. Gavish, M. & Donoho, D. L. Optimal shrinkage of singular values. *IEEE Trans. Inf. Theory* **63**, 2137 (2017).
25. 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).
26. Kalman, R. E. Contributions to the theory of time-optimal control. *Bol. Soc. Mat. Mex.* **5**, 102 (1960).
27. Yan, G. et al. Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* **550**, 519 (2017).
28. Marčenko, V. A. & Pastur, L. A. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik* **1**, 457 (1967).
29. Féral, D. & Péché, S. The largest eigenvalue of rank one deformation of large Wigner matrices. *Commun. Math. Phys.* **272**, 185 (2007).
30. 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).
31. Benaych-Georges, F. & Nadakuditi, R. R. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227**, 494 (2011).
32. 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).
33. Pizzo, A., Renfrew, D. & Soshnikov, A. On finite rank deformations of Wigner matrices. *Ann. I. H. Poincaré PR* **49**, 64–94 (2013).
34. Baik, J., Ben Arous, G. & Péché, S. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33**, 1643 (2005).
35. Valdano, E. & Arenas, A. Exact rank reduction of network models. *Phys. Rev. X* **9**, 031050 (2019).
36. Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F. & Ostojic, S. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Comput.* **33**, 1572 (2021).
37. Gao, P. & Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32**, 148 (2015).
38. Beckermann, B. & Townsend, A. On the singular values of matrices with displacement structure. *SIAM J. Matrix Anal. Appl.* **38**, 1227 (2017).
39. Udell, M. & Townsend, A. Why are big data matrices approximately low rank? *SIAM J. Math. Data Sci.* **1**, 144 (2019).
40. Gao, J., Barzel, B. & Barabási, A.-L. Universal resilience patterns in complex networks. *Nature* **530**, 307 (2016).
41. 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).
42. Jiang, J. et al. Predicting tipping points in mutualistic networks through dimension reduction. *Proc. Natl Acad. Sci. USA* **115**, E639 (2018).
43. Laurence, E., Doyon, N., Dubé, L. J. & Desrosiers, P. Spectral dimension reduction of complex dynamical networks. *Phys. Rev. X* **9**, 011042 (2019).
44. Vegué, M., Thibeault, V., Desrosiers, P. & Allard, A. Dimension reduction of dynamics on modular and heterogeneous directed networks. PNAS Nexus 2, pgad150 (2023).
45. Kundu, P., Kori, H. & Masuda, N. Accuracy of a one-dimensional reduction of dynamical systems on networks. *Phys. Rev. E* **105**, 024305 (2022).
46. 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).
47. Kuehn, C. & Bick, C. A universal route to explosive phenomena. *Sci. Adv.* **7**, 1 (2021).
48. 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).
49. Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1 (2020).
50. Matheny, M. H. et al. Exotic states in a simple network of nanoelectromechanical oscillators. *Science* **363**, 1057 (2019).
51. 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. & Ostojic, 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. Statis.* **46**, 1742 (2018).
69. Scheffer, L. K. et al. A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, 1 (2020).

Methods

Random graphs

A random graph can be described by a random matrix

$$W = \langle W \rangle + R, \tag{8}$$

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,

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

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, spin glasses, 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, $\phi = \phi_{\mathrm{FD}}$, a Fermi–Dirac distribution, and in its weighted version (WDSCM), $\phi = \phi_{\mathrm{BE}}$, 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 $\sigma_1 \geq \ldots \geq \sigma_N$ be the singular values of $\langle W\rangle$. If $\langle W_{ij}\rangle = \phi_{\mathrm{FD}}(L_{ij}) < 1/2$ or $\langle W_{ij}\rangle = \phi_{\mathrm{BE}}(L_{ij})$ for all $i, j \in \{1, \ldots, N\}$, where L is a rank-one matrix, then

$$\sigma_i \leq \sum_{k=i}^{\infty} \ell_k \leq \frac{N\gamma^i}{1-\gamma} \qquad \forall\ i \in \{1, \ldots, N\}, \tag{10}$$

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 $\langle W_{ij}\rangle = \phi_{\mathrm{FD}}(L_{ij}) > 1/2$ 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,i are given by equation (10), which is computed by summing the constants $n_i > n_{i+1} > \ldots$ until n_k is smaller than 10^{-12}. For RPG, the vectors $\mathbf{m}_\mu$ and $\mathbf{n}_\mu$ 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 Λ 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 analysis[72]) but is still subject to new and interesting developments in random matrix theory, data science[24,67] and in network science, which uses hyperbolic geometry[73] and information theory[74]. 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 thrank 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 matrices[32] to determine optimal ways of shrinking the singular values (Supplementary Information section IC). In Fig. 1l, 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: srank ≤ nrank ≤ erank ≤ 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 $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_N \geq 0$ of W_N satisfy the inequalities

$$\alpha^{i-1} \leq \frac{\sigma_i}{\sigma_1} \leq \omega^{i-1}, \qquad i \in \{1, \ldots, N\}. \tag{11}$$

Then, as $N \to \infty$,

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

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, $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_N \geq 0$, satisfy the inequality

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

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 $i \in \{1, \ldots, N\}$. Then,

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

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 $L2$ 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 IIC 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 $M^+ \circ \mathbf{F} \circ M$,

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

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

$$\mathcal{E}(\mathbf{x}) = \|M\mathbf{f}(\mathbf{x}) - \mathbf{F}(M\mathbf{x})\|/\sqrt{n}, \tag{16}$$

where $\|\ \|$ 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

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

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

with $P = M^+M$ being a projector and M^+ being the Moore–Penrose pseudoinverse of M. In principle, the alignment error $\mathcal{E}(\mathbf{x})$ 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

$$\dot{\mathbf{X}} = M\mathbf{f}(M^+\mathbf{X}) \tag{19}$$

is optimal in the sense that it minimizes the alignment error $\varepsilon'(\mathbf{x})$ in $\mathbb{R}^N$. As a consequence, the alignment error $\mathcal{E}'(\mathbf{x})$ 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 $\mathcal{E}(\mathbf{x})$ related to the singular values of W.

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

$$\sqrt{n}\,\mathcal{E}(\mathbf{x}) \le \|V_n^\top J'_{\mathbf{x}}(I - V_nV_n^\top)\mathbf{x}\| + \sigma_{n+1}\|V_n^\top J'_{\mathbf{y}}\|_2\,\|\mathbf{x}\|, \tag{20}$$

where $\mathbf{y}' = W\mathbf{x}'$ with $\mathbf{x}'$ being some point between $\mathbf{x}$ and $V_nV_n^\top\mathbf{x}$, σ_i is the ith singular value of W, and $J'_{\mathbf{x}} = J_{\mathbf{x}}(\mathbf{x}', \mathbf{y}')$ and $J'_{\mathbf{y}} = J_{\mathbf{y}}(\mathbf{x}', \mathbf{y}')$ 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:

$$\frac{\mathcal{E}(\mathbf{x})}{\|\mathbf{x}\|} \le \frac{1}{\sqrt{n}}\left[\alpha(\mathbf{x}', \mathbf{y}') + \sigma_{n+1}\beta(\mathbf{x}', \mathbf{y}')\right], \tag{21}$$

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

As a bonus, the proof of the theorem suggests choosing M as the truncated right singular vectors $V_n^\top$, 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 $J_{\mathbf{x}}(\mathbf{x}', \mathbf{y}') = dI$ for some real constant d and n is the rank of W, then $\mathcal{E}(\mathbf{x}) = 0$ (Supplementary Corollary 74 in Supplementary Information section IIID). For example, consider the class of dynamics in matrix form

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

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

$$\dot{\mathbf{X}} = d\mathbf{X} + V_r^\top\mathbf{s}(U_r\Sigma_r\mathbf{X}), \tag{23}$$

where $\mathbf{X} = V_r^\top\mathbf{x}$. For any n and $\mathbf{X} = V_n^\top\mathbf{x}$, 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

$$\frac{\mathcal{E}(\mathbf{x})}{\|\mathbf{x}\|} \le \frac{\sigma_{n+1}}{\sqrt{n}}, \tag{24}$$

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

IIIC) have the general form $\dot{x}_i = h_i(x_i, y_i)$ for all $i \in \{1, \ldots, N\}$, where $x_i : [0, \infty) \to \mathbb{R}$, $y_i = \sum_{j=1}^{N} W_{ij} x_j$ and $h_i : \mathbb{R}^2 \to \mathbb{R}$ 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

$$\dot{X}_\mu = \mathcal{C}_\mu + \sum_{d_x=1}^{\infty} \sum_{\boldsymbol{\alpha}} \mathcal{D}^{(d_x+1)}_{\mu\boldsymbol{\alpha}} X_{\boldsymbol{\alpha}} + \sum_{d_y=1}^{\infty} \sum_{\boldsymbol{\beta}} \mathcal{W}^{(d_y+1)}_{\mu\boldsymbol{\beta}} X_{\boldsymbol{\beta}} + \sum_{d_x,d_y=1}^{\infty} \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} \mathcal{T}^{(d_x+d_y+1)}_{\mu\boldsymbol{\alpha}\boldsymbol{\beta}} X_{\boldsymbol{\alpha}\boldsymbol{\beta}},$$

where we have introduced the multi-indices $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{d_x})$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{d_y})$ with $\alpha_p, \beta_q \in \{1, \ldots, n\}$, the compact notation for products $X_{\boldsymbol{\gamma}} = X_{\gamma_1} \ldots X_{\gamma_d}$, whereas $\mathcal{C}_\mu$ denotes a real constant and $\mu \in \{1, \ldots, n\}$. 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

$$\mathcal{D}^{(d_x+1)}_{\mu\boldsymbol{\alpha}} = \sum_{i=1}^{N} c_{id_x0} M_{\mu i} M^{+}_{i\boldsymbol{\alpha}},$$

$$\mathcal{W}^{(d_y+1)}_{\mu\boldsymbol{\beta}} = \sum_{i=1}^{N} \sum_{\mathbf{j}} c_{i0d_y} M_{\mu i} W_{i\mathbf{j}} M^{+}_{\mathbf{j}\boldsymbol{\beta}},$$

$$\mathcal{T}^{(d_x+d_y+1)}_{\mu\boldsymbol{\alpha}\boldsymbol{\beta}} = \sum_{i=1}^{N} \sum_{\mathbf{j}} c_{id_xd_y} M_{\mu i} M^{+}_{i\boldsymbol{\alpha}} W_{i\mathbf{j}} M^{+}_{\mathbf{j}\boldsymbol{\beta}},$$

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

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 (rtol = 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 weigh. 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 optht 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.

https://doi.org/10.1038/s41567-023-02303-0

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

*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, γ is the global infection rate and the global observable $\mathcal{X}^* = \mathbf{w} \cdot \mathbf{X}^*$ is defined with $\mathbf{w} = (w_1\, 0 \ldots 0)^\top$ 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, γ is the global microbial interaction weight, and the global observable is defined with $\mathbf{w} = (w_1 \ldots w_{n_{\min}}\, 0 \ldots 0)^\top$ where $\bar{\mathbf{w}} = (w_1\ \ldots\ w_{n_{\min}}) = V_{n_{\min}}^\top \mathbf{1}/(10N)$ 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 $x_i = \mathrm{e}^{\mathrm{i}\theta_i}$, 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 $X_1, \ldots, X_n$ are, thus, complex. The parameters of the RNN in Fig. 4h are d_i=1 for all i and γ=2. The neuronal dynamics is the Wilson–Cowan dynamics, where $\mathcal{S}$ denotes the sigmoid function $x \mapsto 1/(1+\mathrm{e}^{-x})$. Its parameters in Fig. 4f are a=0.05, b=1, c_i=1, d_i=1 for all i, γ is the global synaptic weight and the global observable is defined with $\mathbf{w} = (w_1\, 0 \ldots 0)^\top$ 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.

CONTENTS

I. PRELIMINARIES ON SINGULAR VALUE DECOMPOSITION

Singular Value Decomposition (SVD) goes back to Beltrami (1873) and Jordan (1874) and has become a central linear algebra tool in many areas of science, partly because of its fundamental role in dimension reduction [1–3][4, Chapter 1]. Although one must be careful with the comparisons, which have led to abuses of language [5], SVD possesses some similarities with techniques such as Principal Component Analysis (PCA) [6–12], Karhunen-Loève Transform (KLT) [13–15], Proper Orthogonal Decomposition (POD) [16–18], and Empirical Orthogonal Function (EOF) [19, 20]. In machine learning, some autoencoders have been shown to be at best equivalent to SVD [21, 22]. Even if the subject is old in itself, there are still many interesting developments about SVD, notably in random matrix theory [23–31] where the singular value distribution is often called the eigenvalue distribution of the Wishart, chiral or Laguerre matrix ensembles [32, Chap. 3] or of sample covariance matrices [23, Chap. 3]. Because of its importance in our work and for the sake of completeness, we gather fundamental theorems related to SVD which will be useful to prove the main mathematical results of the paper. We begin this section by recalling the definition of SVD and its close relationship with the rank, i.e., the maximal number of linearly independent rows or columns of a matrix.

A. Definition of SVD and its link to the rank

First of all, any matrix admits a factorization based on its rank. Indeed, if A is a matrix of dimension $m \times n$ and of rank r, then there exists a rank factorization of A, i.e., a decomposition of the form $A = LM$, where L and M are matrices of dimension $m \times r$ and $r \times n$, respectively. Moreover, the rank factorization $A = LM$ is not unique. One very popular rank factorization valid, in particular, for real symmetric matrices is the eigenvalue decomposition. Yet, an arbitrary matrix A is not always diagonalizable by a similarity relation $A = PDP^{-1}$ (e.g., any rectangular matrix). Note, however, that the matrices $AA^\dagger$ and $A^\dagger A$ († denoting the Hermitian conjugation) are square and diagonalizable by a unitary matrix since they are Hermitian (hence, normal). Using this important remark, it can be shown that there always exists a unitary equivalence relation between a matrix and a diagonal matrix of nonnegative elements, the singular value decomposition.

Theorem S1. *Let A be a complex matrix of dimension $m \times n$ and rank r. Then, there exists a SVD of A, i.e., a factorization of the form*

$$A = U\Sigma V^\dagger \tag{S1}$$

where $U = (u_1, ..., u_m)$ and $V = (v_1, ..., v_n)$ are unitary matrices of dimension $m\times m$ and $n\times n$, containing respectively the eigenvectors u_i of $AA^\dagger$ and the eigenvectors v_i of $A^\dagger A$. Moreover, the matrix Σ is a rectangular diagonal matrix of size $m \times n$ defined as

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & ... \\ 0 & \sigma_2 & ... \\ \vdots & \vdots & \ddots \end{pmatrix} \quad \textit{with} \quad \begin{array}{l} \sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0 \\ \sigma_{r+1} = ... = \sigma_q = 0 \end{array} \tag{S2}$$

where $q = \min(m, n)$ and $\sigma_i = \sqrt{\lambda_i}$ with λ_i being the i-th eigenvalue of $A^\dagger A$ or $AA^\dagger$. If additionally all the elements of A are real, then U and V are real orthogonal matrices.

Proof. See theorem 3.1.1 of Ref [33], theorem 2.6.3 of Ref. [34], or theorem 1.3.9 of Ref. [25]. □

Remark S2. The nonnegative numbers $\sigma_1, ..., \sigma_q$ in the previous theorem are called the *singular values* of A while the vectors $u_1, ..., u_m$ and $v_1, ..., v_n$ are respectively called the *left and right singular vectors* of A. For clarity, especially when the singular values of multiple matrices are involved, we will define σ_i as a function of A and write its values as $\sigma_i(A)$.

Remark S3. In general, there is no obvious relationship between the eigenvalues and the singular values of a (square) matrix. However, for the family of normal matrices (including hermitian, anti-hermitian, unitary, and anti-unitary matrices), the singular values are given by the module of the eigenvalues. To visualize the singular values, it is typical to plot them in a decreasing order, which is called a scree plot in the context of PCA [9, 10], or illustrate them in a histogram.

The SVD is thus closely related to the notion of rank, since the number of nonzero singular values of a matrix is *equal* to its rank (while the number of its nonzero eigenvalues is lower or equal to its rank [34, p.151]). Its relation to dimension reduction then becomes obvious: one can truncate the matrices U, V, and Σ by removing their last columns (and rows for Σ) to get smaller matrices $U_r = (u_1 ... u_r)$, $V_r = (v_1 ... v_r)$, and $\Sigma_r = \mathrm{diag}(\sigma_1, ..., \sigma_r)$ with $r = \mathrm{rank}\, A$, and obtain a rank factorization:

$$A = U_r \Sigma_r V_r^\dagger, \tag{S3}$$

which is sometimes called the compact singular value decomposition. More importantly for dimension reduction, when the matrices U, V, and Σ are truncated to U_k, V_k, Σ_k with $k < n$, the truncated SVD is the optimal low-rank factorization as it will be seen in the next subsection.

Remark S4. It is often more convenient to rewrite the SVD in Eq. (S1) or equivalently in Eq. (S3) as

$$A = \sum_{i=1}^{r} \sigma_i \, u_i \, v_i^\dagger \, . \tag{S4}$$

This shows that any matrix of rank r is equal to the sum of r linearly independent unitary matrices, each being of rank 1 and having a (Frobenius or spectral) norm equal to 1. If all the singular values are distinct, then $\sigma_1 \, u_1 \, v_1^\dagger$ and $\sigma_r \, u_r \, v_r^\dagger$ respectively constitute the most and the least important contributions to the matrix A. Moreover, Eq. (S4) implies an explicit formula for the Moore-Penrose pseudo-inverse of A,

$$A^+ = \sum_{i=1}^{r} \frac{1}{\sigma_i} \, v_i \, u_i^\dagger \, , \tag{S5}$$

proving that A and A^+ share the same rank.

B. Weyl's theorem and optimal low-rank factorization

The SVD shares many equivalent theorems with the eigenvalue decomposition [34], such as Rayleigh's theorem, the Courant-Fischer theorem, Cauchy's interlacing theorem, and, in particular, Weyl's theorem, which is of fundamental importance in the paper. The following result was obtained in 1951 by Fan [35, Theorem 2].

Theorem S5. *Let A and B be two matrices of dimension $m \times n$ and let $q = \min(m, n)$. Then,*

$$\sigma_{i+j-1}(A + B) \leq \sigma_i(A) + \sigma_j(B) \qquad \forall\, 1 \leq i,\, j,\, i + j - 1 \leq q, \tag{S6}$$

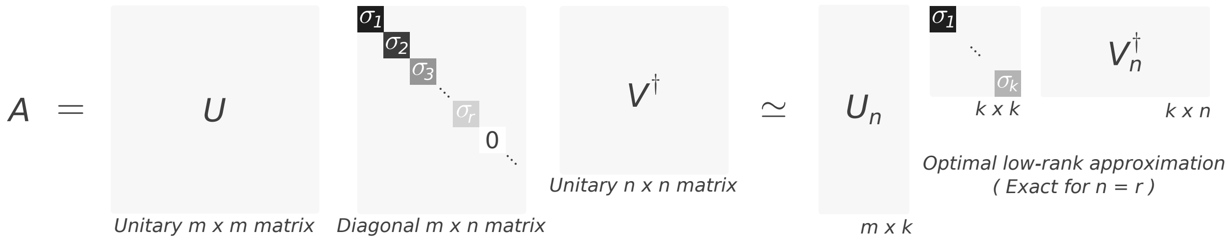

Fig. S1: Truncated SVD is the optimal low-rank approximation of any matrix A according to unitarily invariant norms.

where $\sigma_k(X)$ is the k-th singular value of X and the singular values are ordered in the usual decreasing order.

Proof. A detailed proof based on Weyl's theorem can be done by following the steps of Horn & Johnson [34]. A proof that uses the Courant-Fisher theorem for singular values is also given in Ref. [33, Theorem 3.3.16]. □

Remark S6. If $i = j = 1$, then the previous theorem implies that the dominant singular values satisfy

$$\sigma_1(A+B) \leq \sigma_1(A) + \sigma_1(B). \tag{S7}$$

The latter inequality was known before the generalization by Ky Fan and it is often attributed [36] to Wittmeyer [37, Eq. (VIII)], but Wittmeyer himself writes in a footnote that the equation is in Wintner,"Spektraltheorie der unendlicheri Matrizen", Leipzig 1929, p. 130. Nowadays, the result is, perhaps, not surprising: it is the triangle inequality for the spectral matrix norm.

A first key corollary [25, Exercise 1.3.22 (iv)] allows us to analyze random graphs through perturbation theory of random matrices. Indeed, the following result establishes that the strength (norm) of a matrix perturbation bounds the difference between each singular value of a matrix and the ones of its perturbed version.

Corollary S7. *Let A and B be two matrices of dimension $m \times n$ and let $q = \min(m, n)$.*

$$|\sigma_i(A+B) - \sigma_i(A)| \leq \|B\|_2 \qquad \forall\, 1 \leq i \leq q, \tag{S8}$$

where $\sigma_i(X)$ is the i-th singular value of X and the singular values are ordered in the usual decreasing order.

The importance of the Weyl theorem in the paper also relies on what it implies for dimension reduction. In particular, it allows proving the Schmidt-Eckart-Young-Mirsky theorem [1–3, 38–40] (often called the Eckart-Young theorem [41, Theorem 2.4.8] or the Eckart-Young-Mirsky theorem [42]) which shows that the truncated SVD is the optimal low-rank approximation of a matrix according to unitarily invariant norms. In Theorem S8, we present our formulation of the result (illustrated in Fig. S1) for the Frobenius norm and the spectral norm.

Theorem S8. *Let A be a matrix of rank greater than or equal to k. Consider the optimization problem*

$$\begin{aligned} &\underset{B}{\text{minimize}} \quad \|A - B\|^2 \\ &\text{subject to} \;\; \operatorname{rank} B \leq k\,, \end{aligned} \tag{P0}$$

where $\|\;\|$ denotes the spectral norm $\|\;\|_2$ or the Frobenius norm $\|\;\|_F$. Then, the minimum error of problem (P0) is

$$\min_{B, \operatorname{rank} B \leq k} \|A - B\|_2^2 = \sigma_{k+1}^2 \quad \textit{or} \quad \min_{B, \operatorname{rank} B \leq k} \|A - B\|_F^2 = \sum_{i=k+1}^{q} \sigma_i^2, \tag{S9}$$

where $q = \min\{m, n\}$ and $\sigma_1 \geq \ldots \geq \sigma_q$ are the singular values of A. Furthermore, in both cases, a solution to problem (P0) is provided by the k-truncated SVD of A, i.e.,

$$B^* = \underset{B, \operatorname{rank} B \leq k}{\arg\min} \|A - B\|_2^2 = \underset{B, \operatorname{rank} B \leq k}{\arg\min} \|A - B\|_F^2 = \sum_{i=1}^{k} \sigma_i u_i v_i^\dagger, \tag{S10}$$

where u_i, v_i are the i-th left and right singular vectors of A, respectively. The solution B^* is unique if $\sigma_k > \sigma_{k+1}$.

For our paper, especially to find the upper bound on the alignment error [Theorem S72], Theorem S8 entails another important result: the projectors formed by the left and right singular vector matrices are optimal orthogonal projectors. This fact seems to be well known [43, Fact 2] but, to the authors' knowledge, has not yet been presented in a comprehensive form accompanied by a detailed proof. We hence introduce the following theorem, which will be used later to prove Theorem S72.

Theorem S9. *Let A be a $m \times n$ real matrix of rank r with singular value decomposition $U\Sigma V^\top$ and k-truncated singular value decomposition $U_k\Sigma_k V_k^\top$. Let $\| \|$ denote the spectral norm $\| \|_2$ or the Frobenius norm $\| \|_F$. Consider the optimization problem*

$$\text{minimize } \|(I - M^+M)A\|^2, \tag{P1}$$

where the optimization variable M is a $k \times m$ matrix such that $k \leq m$.

1. *If $k = n \leq m$, then $M = A^+$ solves the problem (P1) with error 0.*
2. *If $k = m$, then any matrix M with rank m solves the problem (P1) with error 0.*
3. *If $k \leq r < \min(m, n)$, then $M = U_k^\top$ minimizes problem (P1) with errors*

$$\min_{M, \text{rank } M \leq k} \|(I - M^+M)A\|_2^2 = \sigma_{k+1}^2 \quad \textit{and} \quad \min_{M, \text{rank } M \leq k} \|(I - M^+M)A\|_F^2 = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2, \tag{S11}$$

which are equal to zero if $k = r$.

Similarly, let B be a $\ell \times m$ real matrix of rank r with singular value decomposition $LSR^\top$ and k-truncated singular value decomposition $L_k S_k R_k^\top$. Consider the optimization problem

$$\text{minimize } \|B(I - M^+M)\|^2 \tag{P2}$$

where, again, the optimization variable M is a $k \times m$ matrix with $k \leq m$.

4. *If $k = \ell \leq m$, then $M = B$ solves the problem (P2) with error 0.*
5. *If $k = m$, then any matrix M with rank m solves the problem (P2) with error 0.*
6. *If $k \leq r < \min(\ell, m)$, then $M = R_k^\top$ minimizes problem (P2) with errors*

$$\min_{M, \text{rank } M \leq k} \|B(I - M^+M)\|_2^2 = \sigma_{k+1}^2 \quad \textit{and} \quad \min_{M, \text{rank } M \leq k} \|B(I - M^+M)\|_F^2 = \sum_{i=k+1}^{\min(\ell,m)} \sigma_i^2, \tag{S12}$$

which are equal to zero if $k = r$.

Proof. We first consider problem (P1) and prove items 1–3.

1. If $k = n \leq m$, then the dimensions of the matrices M and A^+ coincide and one can choose $M = A^+$. Hence

$$\|(I - M^+M)A\|^2 = \|A - AA^+A\|^2 = \|A - A\|^2 = 0,$$

since $(A^+)^+ = A$ and $AA^+A = A$ by the defining properties of the Moore-Penrose pseudo-inverse [44].

2. If $k = m$, then M is square. Any rank m matrix M of dimension $m \times m$ is invertible, so $M^+ = M^{-1}$ and $I - M^{-1}M = 0$, which implies that $\|(I - M^+M)A\|^2 = 0$.

3. We first prove that

$$\min_{M, \text{rank } M \leq k} \|(I - M^+M)A\|^2 = \min_{\substack{C \\ C = M^+MA, \text{rank } M \leq k}} \|A - C\|^2.$$

Indeed, due to Sylvester's rank inequality [34, Section 0.4.5 (c)] and the inequality $\text{rank } M \leq k$,

$$\text{rank}(M^+M) \leq \min\{\text{rank } M, \text{rank } M^+\} \leq k,$$

which in turn implies that

$$\text{rank}(M^+MA) \leq \min\{\text{rank}(M^+M), r\} = \text{rank}(M^+M) \leq k,$$

where the equality follows from $k \leq r$. Thus,

$$\min_{M, \text{rank } M \leq k} \|(I - M^+M)A\|^2 = \min_{\substack{C \\ C = M^+MA, \text{ rank } C \leq k}} \|A - C\|^2. \tag{S13}$$

Let us now focus on the Frobenius norm. The new form of the problem in Eq. (S13) is compatible with Theorem S8, but with the additional equality constraint that $C = M^+MA$, which directly implies the inequality

$$\min_{\substack{C \\ C = M^+MA, \text{ rank } C \leq k}} \|A - C\|^2 \geq \min_{\substack{C \\ \text{rank } C \leq k}} \|A - C\|^2 = \sum_{i=k+1}^{\min\{m,n\}} \sigma_i^2$$

or equivalently, from Eq. (S13),

$$\min_{M,\mathrm{rank}\, M\leq k} \|(I-M^+M)A\|_F^2 \geq \sum_{i=k+1}^{\min\{m,n\}} \sigma_i^2. \tag{S14}$$

Therefore, if we find a matrix M that reaches the lower bound of inequality (S14), then the minimization problem is solved. Below, we prove that $M = U_k^\top$ is such a solution.

The matrix $I - M^+M$ is an orthogonal projector (this is directly proven from the properties of the Moore-Penrose pseudoinverse) and therefore,

$$(I - M^+M)^\top (I - M^+M) = (I - M^+M)^2 = I - M^+M.$$

The cyclic property of the trace and the eigenvalue decomposition of $AA^\top$ from the SVD imply

$$\|(I-M^+M)A\|_F^2 = \mathrm{tr}\left[AA^\top(I-M^+M)\right] = \mathrm{tr}\left[U\Sigma^2U^\top(I-M^+M)\right].$$

Let M be equal to $U_k^\top$. Then, $M^+M = U_kU_k^\top$ and

$$\|(I-M^+M)A\|_F^2 = \mathrm{tr}\left[\sum_{i=1}^{\min(m,n)} \sigma_i^2 u_i u_i^\top - \sum_{i=1}^{\min(m,n)}\sum_{j=1}^{k} \sigma_i^2 u_i u_i^\top u_j u_j^\top\right].$$

Since $u_i^\top u_j = \delta_{ij}$, we obtain

$$\sum_{i=1}^{\min(m,n)}\sum_{j=1}^{k} \sigma_i^2 u_i u_i^\top u_j u_j^\top = \sum_{i=1}^{\min(m,n)}\sum_{j=1}^{k} \sigma_i^2 u_i \delta_{ij} u_j^\top = \sum_{j=1}^{k} \sigma_j^2 u_j u_j^\top$$

and thus,

$$\|(I-M^+M)A\|_F^2 = \mathrm{tr}\left[\sum_{i=k+1}^{\min(m,n)} \sigma_i^2 u_i u_i^\top\right] = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2\, \mathrm{tr}\left[u_i u_i^\top\right] = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2.$$

Hence, $B^* = U_kU_k^+A$ is a solution to the problem (P0). If $k = r$, then $\|(I-M^+M)A\|_F^2 = \sum_{i=r+1}^{\min(m,n)} \sigma_i^2 = 0$, because $\sigma_i = 0$ for all $i > r$.

For the spectral norm, with $M = U_k^\top$, we have $(I - M^+M)A = (I - U_kU_k^\top)U\Sigma V^\top$ which is equal to

$$\sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^\top - \sum_{i=1}^{k} \sigma_i \sum_{j=1}^{k} u_j u_j^\top u_i v_i^\top - \sum_{i=k+1}^{\min(m,n)} \sigma_i \sum_{j=1}^{k} u_j u_j^\top u_i v_i^\top = \sum_{i=k+1}^{\min(m,n)} \sigma_i u_i v_i^\top,$$

where we have used $u_j^\top u_i = \delta_{ij}$ in the last two terms and the fact that i is never equal j in the last term. We conclude that $\|(I-M^+M)A\|_2 = \|\sum_{i=k+1}^{\min(m,n)} \sigma_i u_i v_i^\top\| = \sigma_{k+1}$ which shows that $M = U_k^\top$ minimizes the error in problem (P1).

The proofs of items 4–6 related to problem (P2) closely follow that of items 1–3. □

There is an interesting data science application for Theorem S9 as explained in the following example.

Example S10. Let X be a $m \times T$ data matrix where m is the number of variables (features) and T is the number of time steps (samples). Then, choosing $M = U_n^\top$ where $U_n = (u_1 \; ... \; u_n)$ with u_μ being the μ-th left singular vector of the data matrix X gives the minimal error to the optimization problem (P1) with $A = X$ and $d = T$. This particular example is related to the so-called proper orthogonal decomposition [40, p.278-279].

C. Effective ranks

In this section, we give more details about the different effective ranks presented in Table II.

- The *stable rank* [45, Definition 7.6.7], also called numerical rank [46], is defined as

$$\mathrm{srank}(A) = \frac{\|A\|_F^2}{\|A\|_2^2} = \frac{\sum_{i=1}^r \sigma_i^2}{\sigma_1^2}. \tag{S15}$$

  It thus measures the relative importance of the sum of the squared singular values with respect to the squared largest singular value. More colloquially, srank(A) compares the total energy of A with the energy contained in the first component (first singular vectors) of A. Note that

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{r}\|A\|_2\,, \tag{S16}$$

where $r = \mathrm{rank}(A)$. From the second inequality, we easily deduce the following upper bound:

$$\mathrm{srank}(A) \leq r\,.$$

The stable rank is stable in the sense that it remains essentially unchanged under a small perturbation of the matrix A, contrarily to the rank [46]. It is used in the design of fast (randomized) algorithms for low-rank approximations [43, 47]. Because it also quantifies to what extent the elements of the matrix are gathered around the diagonal, the stable rank also measures the complexity of the connection patterns between the modules of a network [48].

- The *nuclear rank* [49, p.2183] is defined as

$$\mathrm{nrank}(A) = \frac{\|A\|_*}{\|A\|_2} = \frac{\sum_{i=1}^{r}\sigma_i}{\sigma_1}\,, \tag{S17}$$

where $\|\;\|_*$ is the nuclear norm, also known as the trace norm or the Ky Fan norm. Similarly to the stable rank, it measures the relative importance of the sum of the singular values with respect to the largest singular value. The nuclear norm is upper-bounded such that

$$\|A\|_* \leq \sqrt{r}\|A\|_F \leq r\|A\|_2\,, \tag{S18}$$

where we have used the first inequality of Eq. (S16). Therefore, we find that

$$\mathrm{nrank}(A) \leq \sqrt{r}\;\mathrm{srank}^{1/2}(A) \leq r\,. \tag{S19}$$

- The energy ratio, also called the cumulative explained variance, the reconstructed proportion, or the R_v coefficient [10], is

$$E(\ell) = \frac{\|A_\ell\|_F^2}{\|A\|_F^2} = \frac{\sum_{i=1}^{\ell}\sigma_i^2}{\sum_{j=1}^{r}\sigma_j^2}\,, \tag{S20}$$

where A_ℓ is the ℓ-truncated SVD of A. The *energy ratio* effective rank is

$$\mathrm{energy}(A) = \min\left(\underset{\ell\in\{1,\ldots,N\}}{\arg\max}\,\big(E(\ell) > \tau\big)\right), \tag{S21}$$

where $\tau \in (0,1)$ is a threshold to be chosen. Note that this "graph energy" differs (but is related) to the ones introduced in combinatorics by Gutman and Nikiforov that have applications in theoretical chemistry and spectral graph theory [50, 51].

- Let the coordinate (x_i, y_i) of the i-th singular values be given by

$$x_i = \frac{i-1}{N-1} \quad\text{and}\quad y_i = \frac{\sigma_i - \sigma_N}{\sigma_1 - \sigma_N} \tag{S22}$$

for all $i \in \{1, ..., N\}$, such that the largest singular value is at $(0,1)$ and the smallest singular value is at $(1,0)$. The distance between the line $L = \{(x,y)\,|\,x+y=1\}$, passing through the largest and the smallest singular value, and the position (x_i, y_i) of the i-th singular value is

$$d_i = \frac{1}{\sqrt{2}}\;|x_i + y_i - 1|\,. \tag{S23}$$

The *elbow position* is the largest distance between in $\{d_1, ..., d_N\}$, i.e.,

$$i_{\mathrm{elbow}} = \underset{i\in\{1,\ldots,N\}}{\arg\max}\;d_i. \tag{S24}$$

The *elbow rank* is thus defined as the number of singular values above the position of the elbow, which is described by

$$\mathrm{elbow}(A) = i_{\mathrm{elbow}} - 1 = \frac{1}{\sqrt{2}}\underset{i\in\{1,\ldots,N\}}{\arg\max}\left|\frac{i-1}{N-1} + \frac{\sigma_i - \sigma_N}{\sigma_1 - \sigma_N} - 1\right| - 1. \tag{S25}$$

This effective rank is often used as a rule of thumb to truncate the singular value distribution [27, 52]. It is also named the "scree" or elbow test [10] and may be computed in different ways than above [9].

- Roy and Vetterli's effective rank [53] or Cangelosi and Goriely's information dimension [54] is here called *erank*.

It is defined as

$$\mathrm{erank}(A) = \exp\left[H(p_1, ..., p_r)\right] \tag{S26}$$

where $H(p_1, ..., p_r) = -\sum_{i=1}^{r} p_i \log p_i$ is the Shannon (spectral) entropy, measured in nat, as a function of the singular value mass function

$$p_i = \frac{\sigma_i}{\|A\|_*} = \frac{\sigma_i}{\sum_{j=1}^{r} \sigma_j}, \quad \forall i \in \{1, ..., r\}. \tag{S27}$$

Note that the square of the singular values could be used to define the singular value mass function, as in Ref. [55]. Among other interesting properties, the erank satisfies $1 \le \mathrm{erank}(A) \le r$ and it is naturally related to the minimum coefficient rate [56] (see Ref. [53, Sec. 3] for more details). Moreover, the maximum Shannon entropy is reached for a distribution of identical singular values ($\mathrm{erank}(A) = r \le N$). Intuitively, this means that the erank measures the uniformity of the singular value distribution. For example, the erank of the singular values $(1, 1, 1, 1, 1, 0, 0, 0, 0, 0)$ is 5, the erank of $(5, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ is approximately 7.9, and the erank of $(30, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ is approximately 2.8.

- Let $A = A_\ell + R$ where A_ℓ is a (deterministic) matrix of unknown rank ℓ and R is some noise random matrix. Based on Ref. [57, Definition 4.2] and Ref. [27], the *optimal threshold* $\tau^*(A)$ is defined as

$$\tau^*(A) = \arg\min_{\tau} \|A_\ell - \hat{A}(\tau)\|, \tag{S28}$$

where $\hat{A}(\tau)$ is the τ-truncated SVD of A and $\|\cdot\|$ is some matrix norm (e.g., spectral norm, Frobenius norm). Intuitively, the problem of finding $\tau^*(A)$ is the problem of finding the singular values of the rank–ℓ matrix A_ℓ (signal matrix) by removing the "noisy" singular values of A due to γR. When the level of noise is unknown, under some conditions on R, the optimal threshold

$$\tau^*(A) = \frac{4\sigma_{\mathrm{med}}}{\sqrt{3\,\mu_{\mathrm{med}}}}, \tag{S29}$$

minimizes the Frobenius norm $\|A_\ell - \hat{A}(\tau)\|_F$ in the limit of infinite matrices [27, Corollary 3 and Theorem 1], where σ_{med} is the median of the observed singular value distribution of the weight matrix A and μ_{med} is the median of a Marčenko-Pastur probability density function. The median μ_{med} is generally unknown, but can be computed as explained in Ref. [27, p.5046]. These results, based on random matrix theory [24], are all rigorous in an asymptotic framework under specific conditions given in Ref. [27]. We define thrank(A) has the number of singular values above the optimal singular value threshold $\tau^*(A)$, i.e.,

$$\mathrm{thrank}(A) = \#\big\{\sigma_i \,\big|\, i \in \{1, ..., N\} \text{ and } \sigma_i > \tau^*(A)\big\}, \tag{S30}$$

where # is the cardinal of a set.

- In a similar spirit as the optimal threshold, one can consider the optimal shrinkage of singular values [29, 52, 58] to define an effective rank. Let $A = A_\ell + R$ where A_ℓ is a (deterministic) matrix of unknown rank ℓ and R is some noise random matrix. Shortly, given the singular values of A, the scalar function $s : [0, \infty) \to [0, \infty),\ \sigma_i \mapsto s(\sigma_i)$ is called a *shrinker* or a *denoiser* of singular values. From Refs. [29, 31, 52, 58], one can find analytically the optimal denoiser s^* that minimizes different errors defined from the Frobenius norm, the spectral (operator) norm, or the nuclear norm. We define *shrank* has the rank of the matrix with optimally shrinked singular values, i.e.,

$$\mathrm{shrank}(A) = \#\big\{s^*(\sigma_i) \,\big|\, i \in \{1, ..., N\} \text{ and } s^*(\sigma_i) > 0\,\big\}. \tag{S31}$$

Note that this effective rank also depends on the median of the Marčenko-Pastur distribution when the level of noise is unknown and estimated as in Ref. [29].

Remark S11.

- A simple criterion to determine whether a matrix is low rank can be formulated in terms of the minimal number of elements that are needed to fully describe the matrix by a rank decomposition. More precisely, a $N \times N$ matrix of rank r can be defined to be of low rank if $2rN < N^2$ or identically, $r < N/2$. Similarly, a $N \times N$ matrix of effective rank e can be defined to be of low effective rank if $e < N/2$. However, in the paper, we do not set one criterion to say that a matrix has a low (effective) rank: we rather compare different effective ranks of a graph with the actual rank and dimension of the corresponding matrix.
- To compute the effective ranks exactly, the complete set of singular values is needed, which might not be possible to have for very large matrices (networks). Even if we did not use it in the paper, we acknowledge the fact that the singular values can be approximately obtained by using randomized SVD [59] (e.g., with sklearn.utils.extmath.randomized_svd in *Python*). Note also that the rank is computed from the singular values with a numerical tolerance of 10^{-13} in the paper.

- Among the all the effective ranks listed above, erank, nrank, and srank distinguish themselves by their simple analytic formulation and their clear upper bounds. As we show below, they also enjoy a natural ordering.

Lemma S12 (Effective ranks ordering). *For any matrix A,*

$$\text{srank}(A) \leq \text{nrank}(A) \leq \text{erank}(A) \leq \text{rank}(A)\,. \tag{S32}$$

Proof. Let $r = \text{rank}(A)$. Then A has exactly r positive singular values: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$.

To prove the first inequality, we recall that $\text{srank}(A) = \sum_{i=1}^r (\sigma_i/\sigma_1)^2$ while $\text{nrank}(A) = \sum_{i=1}^r \sigma_i/\sigma_1$. Now, $\sigma_i/\sigma_1 \leq 1$ for all i, so $(\sigma_i/\sigma_1)^2 \leq \sigma_i/\sigma_1$ for all i. Therefore, $\text{srank}(A) \leq \text{nrank}(A)$.

The second and third inequalities both involve $\text{erank}(A)$, which is defined as e^H, where H is the entropy of the probability vector associated with the singular values of A, that is

$$H = \sum_{i=1}^r p_i \ln \frac{1}{p_i}, \qquad p_i = \frac{\sigma_i}{\sum_{j=1}^r \sigma_j}.$$

Going back to the definition of nrank allows us to write

$$p_i = \frac{\sigma_i}{\sigma_1 \sum_{j=1}^r \frac{\sigma_j}{\sigma_1}} = \frac{\sigma_i}{\sigma_1 \text{nrank}(A)}.$$

Hence,

$$H = \sum_{i=1}^r \frac{\sigma_i}{\sigma_1 \text{nrank}(A)} \ln\left(\frac{\sigma_1 \text{nrank}(A)}{\sigma_i}\right) = \ln\big(\text{nrank}(A)\big) + \frac{1}{\text{nrank}(A)} \sum_{i=1}^r \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i}.$$

Therefore,

$$\text{erank}(A) = e^H = \text{nrank}(A) \cdot \exp\left(\frac{1}{\text{nrank}(A)} \sum_{i=1}^r \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i}\right) \geq \text{nrank}(A), \tag{S33}$$

as expected. Note that the equality holds only when $r = 1$. Finally, using the concavity of the logarithm and Jensen's inequality, we deduce that

$$H = \sum_{i=1}^r p_i \ln \frac{1}{p_i} \leq \ln\left(\sum_{i=1}^r p_i \frac{1}{p_i}\right) = \ln(r),$$

which readily implies the last inequality of the lemma. □

Remark S13. There is another type of effective rank, with the same form as the erank, that is related to the srank rather than the nrank (see the proof of the latter lemma). Indeed, we can define the "stable erank" as $\text{serank}(A) = e^H$ with $p_i = \sigma_i^2/(\sum_{j=1}^r \sigma_j^2) = \sigma_i^2/(\sigma_1^2 \text{srank}(A))$. However, we do not use this effective rank in the paper.

Regarding the optimal threshold and shrinkage, we also make the following remarks.

Remark S14.

- In the definition of the optimal threshold and shrinkage, it is assumed that the rank of the signal matrix is finite in the limit $N \to \infty$.
- If the type of noise is unknown, the assumptions of Refs. [27, 29] do not necessarily hold and it is not guaranteed that the threshold and the shrinkage effective ranks are optimal.
- In the GitHub repository low-rank-hypothesis-complex-systems, module singular_values/optimal_shrinkage.py, we provide a Python translation (the first to our knowledge) of the Matlab script optimal_shrinkage.m from Ref.[29]. Moreover, we correct an error made in Ref.[29] concerning the optimal singular value shrinkage for operator norm loss with the Theorem 3.1 of W. Leeb [58]. We also merge and adapt for our purpose the Github repository optht, which is a Python implementation of the Matlab script optimal_SVHT_coef.m [27]. Note that, when applied to a data matrix with unknown noise and a median smaller than the numerical zero (set to 1e-13), the optimal threshold and the optimal shrinkage effective ranks are computed for the singular values greater than 1e-13 only to ensure that the estimated noise is not zero.

We believe that the techniques that lead to the optimal threshold and shrinkage [24, 27, 29, 52, 58] will have a considerable impact on network science (besides, it already has an impact in neuroscience [60]). Indeed, considering noisy networks is a long-standing challenge in network science (e.g., in sociology [61]) that has been addressed, for instance, with Bayesian inference [62–65]. We think that there is still plenty of work to do to denoise or evaluate the level of noise of a network from its singular values. In particular, it would be interesting to find optimal singular value shrinkage functions [29] with noise types that are more specific to real networks and random graphs.

All in all, we have gathered some important results on SVD. We will show how these results can be leveraged in network science, spectral graph theory, and dynamical systems.

II. SVD IN THE STUDY OF COMPLEX SYSTEMS

In this section, we present applications of SVD in the study of complex systems. First, we highlight the ubiquity of the low-rank hypothesis in random graph theory. Second, we present original theorems for the rapid decrease of the singular values in the directed soft configuration model and its weighted version. Third, inequalities and scaling behaviors for the effective ranks are deduced from different decreasing behaviors of the singular values. Fourth, we recall how SVD yields centrality measures for directed networks. Fifth, we discuss about preliminary results concerning the evolution of the effective rank in adaptive systems. Finally, we give a short overview of the use of SVD in dynamical systems.

A. SVD of random graphs

Random graphs and their eigenvalue spectrum have a long and rich history [51, 66–77], but less attention has been given to their singular value decomposition. Indeed, SVD is not mentioned in many of the main introductory textbooks of network science [78] or spectral graph theory [51, 68, 79]. This phenomenon is somewhat expected, since both fields needed to develop their own set of tools, but we believe that SVD deserve much more attention. In the following, we present the low-rank formulation in a wide variety of models ranging from network science and random matrix theory to machine learning and neuroscience.

The adjacency or the weight matrix of a random network model can always be written as

$$W = \langle W \rangle + R\,, \tag{S34}$$

where R is a zero mean random matrix and $\langle W \rangle$ is the (deterministic) expected weight matrix. Typically, $\langle W \rangle$ depends upon a low-rank matrix L:

$$\langle W \rangle = \Phi(L)\,, \tag{S35}$$

where Φ a matrix-valued function of a matrix variable. In all the cases studied below, the (i,j) element $\Phi(L)$ is equal to $\phi(L_{ij})$, with ϕ being a real scalar function of a real variable. To expose the low-rank formulation of $\langle W \rangle$, recall that there always exists a rank factorization

$$\langle W \rangle = LR^\top\,, \tag{S36}$$

where L, R are $N \times r$ matrices and r is the rank of $\langle W \rangle$. Another convenient form is the sum of rank-one matrices

$$\langle W \rangle = \sum_{\mu,\nu=1}^{s} \alpha_{\mu\nu}\boldsymbol{a}_\mu \boldsymbol{c}_\nu^\top = \sum_{\mu=1}^{s} \boldsymbol{a}_\mu \boldsymbol{b}_\mu^\top\,, \tag{S37}$$

where $\boldsymbol{a}_\mu$, $\boldsymbol{c}_\mu$ are $N \times 1$ vectors, $\alpha_{\mu\nu}$ is a real constant for all μ, ν, and $\boldsymbol{b}_\mu = \sum_{\nu=1}^{s} \alpha_{\mu\nu}\boldsymbol{c}_\nu$. Indeed, defining the $N \times s$ matrices $A = (\boldsymbol{a}_1, ..., \boldsymbol{a}_s)^\top$ and $B = (\boldsymbol{b}_1, ..., \boldsymbol{b}_s)^\top$ yields $\langle W \rangle = AB^\top$ and ensures that the rank of W is at most s. In the next examples, we provide the details about the random graphs of Table I in the Methods.

Example S15 (Network science—unweighted graphs)**.** A large class of binary random graphs are described by Bernouilli random matrices, R_{ij} being equal to either $-\langle W_{ij} \rangle$ or $1 - \langle W_{ij} \rangle$. The expected adjacency matrix for...

- ...the $\mathcal{G}(N,p)$ *model* [80–82] with self-loops is

$$\langle W \rangle = L = Np\,\hat{\mathbf{1}}\hat{\mathbf{1}}^\top\,, \tag{S38}$$

  where $Np\hat{\mathbf{1}}\hat{\mathbf{1}}^\top$ is the (exact) SVD of the mean adjacency matrix, which is a rank-one matrix with singular value Np and $N \times 1$ singular vectors $\hat{\mathbf{1}} = (1\,...\,1)^\top/\sqrt{N}$. The model is also called Poisson random graph, Erdős-Rényi model [83], Bernouilli random matrix [84], or spiked Wigner matrix [85].
- ...the *stochastic block model* (SBM) [86, 87] with q communities (generalization of $\mathcal{G}(N,p)$) is

$$\langle W \rangle = L = \sum_{\mu,\nu=1}^{q} \sqrt{n_\mu n_\nu}\, p_{\mu\nu}\, \boldsymbol{b}_\mu \boldsymbol{b}_\nu^\top\,, \tag{S39}$$

  where $p_{\mu\nu}$ is the probability for a vertex in the μ-th block of size n_μ to be connected to a vertex in the ν-th block of size n_ν and b_μ is a block vector with $1/\sqrt{n_\mu}$ at the indices of the μ-th block and zeros elsewhere.

- ... the *Chung-Lu model* [88, 89] is

$$\langle W \rangle = L = \frac{\|\boldsymbol{\kappa}\|^2}{2M} \hat{\boldsymbol{\kappa}} \hat{\boldsymbol{\kappa}}^\top , \tag{S40}$$

where κ is a vector of expected degrees. Note that the annealed approximation, omnipresent in epidemiology [90] (or for spin models [91]), is thus a very strong low-rank hypothesis.

- ... the *metadegree model* [92] is

$$\langle W \rangle = L = \sum_{\mu,\nu=1}^{r} \Delta_{\mu\nu} \, \boldsymbol{v}_\mu \boldsymbol{v}_\nu^\top , \tag{S41}$$

where $(v_\mu)_{\mu=1}^r$ are the N-dimensional vectors of metadegree and Δ is a $r \times r$ nonsingular matrix that contains the "coefficients of mixing" among metadegrees. In Ref. [92, p.2, 2nd column, 2nd paragraph], a low-rank hypothesis is explicitly made as they assume that the rank of $V \Delta V^\top$ is much smaller than the size of the system. One must say, however, that the model is flexible and generalizes, in particular, the Chung-Lu model. The framework developed by the authors of Ref. [92] offers a solid discussion and a strong theoretical ground to better understand, classify, and design random graphs in the future. Their work has inspired our preliminary thoughts on the low-rank hypothesis.

- ... the directed S^1 model of *random geometric networks* [93, 94] has elements

$$\langle W_{ij} \rangle = \phi(L_{ij}) = \frac{1}{1 + L_{ij}^{\beta/2}} , \tag{S42}$$

where $\beta > 0$ (inverse temperature of the Fermi-Dirac distribution). The elements of the matrix L are defined as

$$L_{ij} = \frac{R^2 \theta_{ij}^2}{\mu^2 (\boldsymbol{\kappa}_{\text{in}})_i^2 (\boldsymbol{\kappa}_{\text{out}})_j^2} , \tag{S43}$$

where μ and R are positive constants, the latter representing the radius of the circle on which the vertices are distributed, θ_{ij} is the angular distance between the vertices i and j on the circle, and $(\boldsymbol{\kappa}_{\text{in}})_i, (\boldsymbol{\kappa}_{\text{out}})_i$ denote the i-th latent in- and out-degrees respectively (positive constants). To estimate the rank of the matrix L, it is more convenient to rewrite it using the Hadamard product:

$$L = \frac{R^2}{\mu^2} \left(\bar{\boldsymbol{\kappa}}_{\text{in}} \bar{\boldsymbol{\kappa}}_{\text{out}}^\top \right) \circ \bar{\theta} , \tag{S44}$$

where

$$\bar{\boldsymbol{\kappa}}_{\text{in}} = (\, 1/(\boldsymbol{\kappa}_{\text{in}})_i^2 \,)_{i=1}^N , \qquad \bar{\boldsymbol{\kappa}}_{\text{out}} = (\, 1/(\boldsymbol{\kappa}_{\text{out}})_i^2 \,)_{i=1}^N , \qquad \bar{\theta} = (\, \theta_{ij}^2 \,)_{i,j=1}^N . \tag{S45}$$

Clearly, $\bar{\boldsymbol{\kappa}}_{\text{in}} \bar{\boldsymbol{\kappa}}_{\text{out}}^\top$ is a rank-one matrix. Now, according to Ref. [95, Theorem 7], the rank of a distance matrix (with squared elements) is D, $D+1$, or $D+2$, where D is the dimension of the manifold where the points are embedded. Here $D = 1$, which means that the rank of $\bar{\theta}$ is at most 3. Recalling the well-known inequality $\text{rank}(A \circ B) \leq \text{rank}(A)\,\text{rank}(B)$, we conclude that the matrix L defining the expected adjacency matrix of the S^1 model has a rank of at most 3. For the S^D model [48], one can proceed similarly to conclude that its expected adjacency matrix has a rank of at most $D+2$.

- ... the *soft directed configuration model* also has elements following a Fermi-Dirac distribution such that

$$\langle W_{ij} \rangle = \phi(L_{ij}) = \frac{L_{ij}}{1 + L_{ij}} , \qquad L_{ij} = \alpha_i \beta_j \tag{S46}$$

for some positive parameters α_i, β_j. Thus, $L = \boldsymbol{\alpha}\boldsymbol{\beta}^T$ is a rank-one matrix and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are positive vectors defined in subsection II B.

- ... of the *Barabási-Albert model* (BA) [96], a model of preferential attachment and a particular case of Price's model [97], does not possess an explicit formula of the form $\langle W \rangle = \Phi(L)$. However, in Fig. S2b, we show the singular values of the model for increasing values of m, the number of edges to which a new vertex is attached, and we observe rapid decreases.

Example S16 (Network science—weighted graphs)**.** All the above-mentioned unweighted graph models can be generalized to include weights.

- The simplest procedure consists in posing

$$W = A \circ \mathcal{W} , \tag{S47}$$

where A and $\mathcal{W}$ are two independent $N \times N$ random matrices. The first matrix, A, corresponds to the Bernoulli matrices introduced in Example S15 that control the existence of edges, while $\mathcal{W}$ is a (possibility continuous) random matrix that only encodes the values of the weights. Due to the independence of A and $\mathcal{W}$, the expected weight matrix factorizes as

$$\langle W \rangle = \langle A \rangle \circ \langle \mathcal{W} \rangle \,. \tag{S48}$$

For instance, supposing that the elements of A are i.i.d. such that $\mathrm{Prob}(A_{ij} = 1) = p$, we get the *weighted* $\mathcal{G}(N, p)$ *model* satisfying

$$\mathrm{Prob}(W_{ij} = 0) = 1 - p, \qquad \mathrm{Prob}(W_{ij} \neq 0) = p, \qquad \langle W \rangle = p\langle \mathcal{W} \rangle, \tag{S49}$$

meaning that the average matrix of weights completely determine the rank of $\langle W \rangle$. If we additionally impose that all W_{ij}'s are i.d.d. with mean w, then we conclude that

$$\langle W \rangle = Npw\, \hat{\mathbf{1}}\hat{\mathbf{1}}^\top \,, \tag{S50}$$

thus corresponding to a rank-one model that we denote $\mathcal{G}(N, p, w)$ in Table I.

- When $p = 1$ in Eq. (S49), all edges exist and one recovers the models of complete weighted graphs, such as the simplest form of the *weighted stochastic block model* (WSBM) with q communities (groups) [98, Eq. (2.3)] (where $q = K$), for which the probability density function of the weights is

$$f_W(w) = \prod_{i,j=1}^{N} \frac{1}{\sqrt{2\pi\Sigma_{ij}^2}} e^{-\frac{(w_{ij} - M_{ij})^2}{2\Sigma_{i,j}^2}} \,, \tag{S51}$$

where $M = (M_{ij})_{i,j=1}^N$ and $\Sigma = (\Sigma_{ij})_{i,j=1}^N$ are block matrices whose elements can only take a few different values encoded in the $q \times q$ matrices μ and σ as

$$M_{ij} = \mu_{g_i g_j} \qquad \text{and} \qquad \Sigma_{ij} = \sigma_{g_i g_j} \,. \tag{S52}$$

In the last equations, $g_i \in \{1, \ldots, q\}$ is the group label of vertex i. The expected weight matrix of this WSBM is simply

$$\langle W \rangle = L = M = \sum_{\kappa,\nu=1}^{q} \sqrt{n_\kappa n_\nu}\, \mu_{\kappa\nu}\, \boldsymbol{b}_\kappa \boldsymbol{b}_\nu^\top \,, \tag{S53}$$

where n_κ and $\boldsymbol{b}_\kappa$ are defined as in Example S15 for the SBM. Since the vertices belonging to the same group produce the same rows in matrix M, we conclude that the rank of L is at most q.

- Random models as in Eq. (S47) were also used to define weighted versions of the SBM [99] and planted-partition model [100]. Moreover, they served to study synchronization in random weighted directed networks [101] as well as the spectral properties of neuronal networks with inhibition [102] and the transitions to chaos of dilute random neuronal networks [103]. Separating edges and their weights as in Eq. (S47) is also common practice in random matrix theory when studying the spectral properties of random weighted directed graphs [104–108].
- Although variables A and $\mathcal{W}$ are independent in Eq. (S47), variables A and $W = A \circ \mathcal{W}$ depend on each other. Indeed, they have a non-factorizable joint probability density function (pdf) of the form

$$F_{A,W}(a, w) = \prod_{1 \leq i < j \leq N} \Big(p_{ij}\delta(a_{ij} - 1) f_{\mathcal{W}_{ij}}(w_{ij}) + (1 - p_{ij})\delta(a_{ij})\delta(w_{ij}) \Big) \,, \tag{S54}$$

where δ denotes Dirac's delta distribution, p_{ij} is the marginal probability for $A_{ij} = 1$, and $f_{\mathcal{W}_{ij}}$ is the pdf of the independent variable $\mathcal{W}_{ij}$. Note that we have assumed that the graph is undirected for simplicity. We see from the previous equation that $\mathcal{W}_{ij}$ can interpreted as the random variable W_{ij} given the existence of an edge from j to i, i.e., $W_{ij} \,|\, A_{ij} = 1$. Setting $\mu_{ij} = \langle \mathcal{W}_{ij} \rangle = \int w f_{\mathcal{W}_{ij}}(w) dw$ and returning to Eq. (S48), we conclude that any model defined using Eq. (S54) also satisfies

$$\langle W_{ij} \rangle = p_{ij}\mu_{ij} \,. \tag{S55}$$

A good example of such a model is the S^1 version of the *weighted random geometric model* (WRGM) [109] with fixed hidden variables

$$\boldsymbol{\theta} \in [0, 2\pi)^N, \qquad \boldsymbol{\kappa} \in \mathbb{R}_+^N, \qquad \boldsymbol{\sigma} \in \mathbb{R}_+^N, \tag{S56}$$

and parameters

$$\alpha \in [0, 1), \qquad \beta > 1, \qquad \mu > 0, \qquad \nu > 0, \qquad R > 0 \,. \tag{S57}$$

One can prove that the pdf of this model is given by Eq. (S54) with

$$p_{ij} = \frac{1}{1 + L_{ij}^{\beta}}, \qquad L_{ij} = \frac{R\theta_{ij}}{\mu\kappa_i\kappa_j}, \qquad f_{\mathcal{W}_{ij}}(w) = \frac{\nu\sigma_i\sigma_j}{(\kappa_i\kappa_j)^{1-\alpha}(R\theta_{ij})^{\alpha}} f(\epsilon)\,, \tag{S58}$$

where R stands for the radius of the circle on which the vertices are distributed, θ_{ij} is the angular distance between the vertices i and j, respectively placed at angle θ_i and θ_j on the circle, and ϵ is an auxiliary random variable whose pdf is f and whose mean is equal to 1. The expected weighted matrix is thus

$$\langle\, W_{ij}\,\rangle = \phi(L_{ij}, M_{ij}) = \frac{1}{L_{ij}^{\alpha}(1 + L_{ij}^{\beta})}\, M_{ij}, \qquad M_{ij} = \frac{\nu\sigma_i\sigma_j}{\mu^{\alpha}\kappa_i\kappa_j}\,. \tag{S59}$$

The rank of the corresponding matrices L and M are at most 3 and equal to 1, respectively.

- All multigraphs can be interpreted as weighted graphs in which the weights can only take nonnegative integer values. One of the best known and most widely used examples of a random multigraph, and thus a weighted random graph, is the *degree-corrected stochastic block model* (DCSBM) [110, 111] with q communities whose probability mass function is defined as

$$\mathrm{Prob}(W = w) = \prod_{1\le i,j\le N} \frac{e^{-\alpha_{ij}}\alpha_{ij}^{w_{ij}}}{w_{ij}!}\,, \tag{S60}$$

that is, the random variable W_{ij} follows a Poisson distribution with values $w_{ij} \in \mathbb{N}$ for all i, j and parameter

$$\alpha_{ij} = \lambda_{g_i g_j}(\hat{\boldsymbol{\kappa}}_{\mathrm{in}})_i(\hat{\boldsymbol{\kappa}}_{\mathrm{out}})_j\,. \tag{S61}$$

In the last equation, $g_i \in \{1, \ldots, q\}$ denotes the group (community) label to which vertex i belongs and $\lambda_{g_i g_j}$ is the expected number of edges from group g_j to g_i. Moreover, $(\hat{\boldsymbol{\kappa}}_{\mathrm{in}})_i$ and $(\hat{\boldsymbol{\kappa}}_{\mathrm{out}})_i$ are group-normalized expected in- and out-degrees, i.e.,

$$\sum_{i=1}^{N}(\hat{\boldsymbol{\kappa}}_{\mathrm{in}})_i\delta_{g_i,\,\mu} = 1\,, \qquad \sum_{i=1}^{N}(\hat{\boldsymbol{\kappa}}_{\mathrm{out}})_i\delta_{g_i,\,\mu} = 1 \tag{S62}$$

for all $\mu \in \{1, \ldots, q\}$. The expected weight matrix is thus

$$\langle W\rangle = L = \Lambda \circ (\hat{\boldsymbol{\kappa}}_{\mathrm{in}}\hat{\boldsymbol{\kappa}}_{\mathrm{out}}^{\top}), \tag{S63}$$

where $\Lambda = (\lambda_{g_i g_j})_{i,j=1}^{N}$ is a block matrix of rank at most q while $\hat{\boldsymbol{\kappa}}$ is the vector of expected degrees. Now, according to the rank inequality for the Hadamard product,

$$\mathrm{rank}(L) = \mathrm{rank}\left(\Lambda \circ (\hat{\boldsymbol{\kappa}}_{\mathrm{in}}\hat{\boldsymbol{\kappa}}_{\mathrm{out}}^{\top})\right) \le \underbrace{\mathrm{rank}\,(\Lambda)}_{\le q}\,\underbrace{\mathrm{rank}\left(\hat{\boldsymbol{\kappa}}_{\mathrm{in}}\hat{\boldsymbol{\kappa}}_{\mathrm{out}}^{\top}\right)}_{1} = q. \tag{S64}$$

Thus, including a "rank-one correction" for better describing vertex degrees in SBM does not affect its low-rank property.

- ...the *random dot product graph* [112] is

$$\langle W\rangle = L = XX^{\top}\,, \tag{S65}$$

where X is a $N \times d$ matrix where the rows are the latent positions of each vertex of the graph. The rank of $\langle W\rangle$ is obviously less or equal to d and from Ref. [112, p.14]: "[...]in the RDPG case, where $\boldsymbol{P}$ [(i.e., $\langle W\rangle$)] is of low rank [...]". The model generalizes the SBM, the degree-corrected SBM and the mixed-membership SBM as shown in Theorem 15 of Ref. [112], which is why we classify it in the weighted networks.

- The authors [113, 114] introduced yet another simple method for providing weights to many classical random network models. Using entropy maximization, they defined exponential families of random weighted graphs, such as those whose weights w_{ij} are nonnegative integers and whose probability mass function is

$$\mathrm{Prob}(W = w) = \prod_{1\le i<j\le N}(1 - y_iy_j)(y_iy_j)^{w_{ij}} \tag{S66}$$

where y_i and y_j satisfy $0 < y_iy_j < 1$ for all i, j and are related in a nonlinear way to the expected strengths while controlling the probability of having an edge from j to i. Indeed,

$$\mathrm{Prob}(W_{i,j} = 0) = 1 - y_iy_j \qquad \text{and} \qquad \mathrm{Prob}(W_{i,j} \neq 0) = y_iy_j\,. \tag{S67}$$

We see that the random variable $W_{i,j} + 1$ follows a geometric distribution of parameter $1 - y_iy_j$. One easily shows that the elements of the expected weight matrix follow the Bose-Einstein distribution

$$\langle W_{ij} \rangle = \phi(L_{ij}) = \frac{L_{ij}}{1 - L_{ij}}\,, \qquad L_{ij} = y_i y_j, \tag{S68}$$

meaning that the expected weight matrix $\langle W \rangle$ is a function of the rank-one matrix

$$L = \boldsymbol{y}\boldsymbol{y}^\top\,. \tag{S69}$$

The above random graph is somewhat analogous to the model presented in Eq. (S46). For this reason, we call the weighted graph satisfying Eq. (S66) the *weighted soft configuration model* (WSCM).

- The weighted random graph defined in Eq. (S66) can be related to the Chung-Lu model. Indeed, supposing $y_i y_j \ll 1$, we get

$$\langle W_{ij} \rangle = y_i y_j + O\big((y_i y_j)^2\big)\,. \tag{S70}$$

Comparing with Eq. (S40), we conclude that the limit case $\langle W \rangle = \boldsymbol{y}\boldsymbol{y}^\top$, where the rank of the expected weight matrix is exactly equal to one, corresponds to a *weighted Chung-Lu model*, sometimes called weighted configuration model (cf. [115, Eq. 5]).

- To define the *weighted directed soft configuration model* (WDSCM), we need to modify Eq. (S66). First, we introduce new positive parameters, say $\bar{y}_1, \ldots, \bar{y}_N$. Then we set

$$\mathrm{Prob}(W = w) = \prod_{1 \le i,j \le N} (1 - y_i \bar{y}_j)(y_i \bar{y}_j)^{w_{ij}}, \qquad w_{ij} \in \mathbb{N}, \tag{S71}$$

which means that $W_{ij} + 1$ a random variable having a geometric distribution of parameter $1 - y_i \bar{y}_j$. Therefore,

$$\langle W_{ij} \rangle = \phi(L_{ij}) = \frac{L_{ij}}{1 - L_{ij}}\,, \qquad L_{ij} = y_i \bar{y}_j,\,, \tag{S72}$$

meaning that the edge directionality has no impact on the rank of the expected weight matrix.

- The model defined by Eq. (S66) is a special case of the following random graph, also introduced in [113]:

$$\mathrm{Prob}(W = w) = \prod_{1 \le i < j \le N} \frac{(x_i x_j)^{\Theta(w_{ij})}(1 - y_i y_j)(y_i y_j)^{w_{ij}}}{1 - y_i y_j + x_i x_j y_i y_j}, \qquad w_{ij} \in \mathbb{N}, \tag{S73}$$

where Θ denotes the Heaviside function and the x_i's are positive parameters. We call the corresponding random graph the *general weighted soft configuration model* (GWSCM). For this general model, the expected weight matrix depends upon two rank-one matrices, L and M, and its elements follow a Bose-Fermi distribution [113]. Explicitly,

$$\langle\, W_{ij} \,\rangle = \phi(L_{ij}, M_{ij})\,, \tag{S74}$$

where

$$\phi(L_{ij}, M_{ij}) = \frac{L_{ij} M_{ij}}{(1 - L_{ij} + L_{ij} M_{ij})(1 - L_{ij})}\,, \qquad L_{ij} = y_i y_j, \qquad M_{ij} = x_i x_j\,. \tag{S75}$$

- Eqs. (S73) can be easily generalized to include directed edges as follows:

$$\mathrm{Prob}(W = w) = \prod_{1 \le i,j \le N} \frac{(x_i \bar{x}_j)^{\Theta(w_{ij})}(1 - y_i \bar{y}_j)(y_i \bar{y}_j)^{w_{ij}}}{1 - y_i \bar{y}_j + x_i \bar{x}_j y_i \bar{y}_j}, \qquad w_{ij} \in \mathbb{N}, \tag{S76}$$

where all parameters are positive. We call the random graph whose weight matrix satisfies (S76) the *general weighted directed soft configuration model* (GWDSCM). Its expected weight matrix is

$$\langle\, W_{ij} \,\rangle = \phi(L_{ij}, M_{ij})\,, \tag{S77}$$

where

$$\phi(L_{ij}, M_{ij}) = \frac{L_{ij} M_{ij}}{(1 - L_{ij} + L_{ij} M_{ij})(1 - L_{ij})}\,, \qquad L_{ij} = y_i \bar{y}_j, \qquad M_{ij} = x_i \bar{x}_j\,. \tag{S78}$$

There are also counter-examples of network model with high effective ranks. The most obvious examples are perhaps the Watts-Strogatz model, which had a considerable impact in the network science, and some non-random graphs.

Example S17. The *Watts-Strogatz model* [116] is described by the random matrix

$$W = D_k + R \tag{S79}$$

where D_k is a band matrix of bandwidth k whose k up- and sub- diagonal entries are equal to 1 while R is a matrix

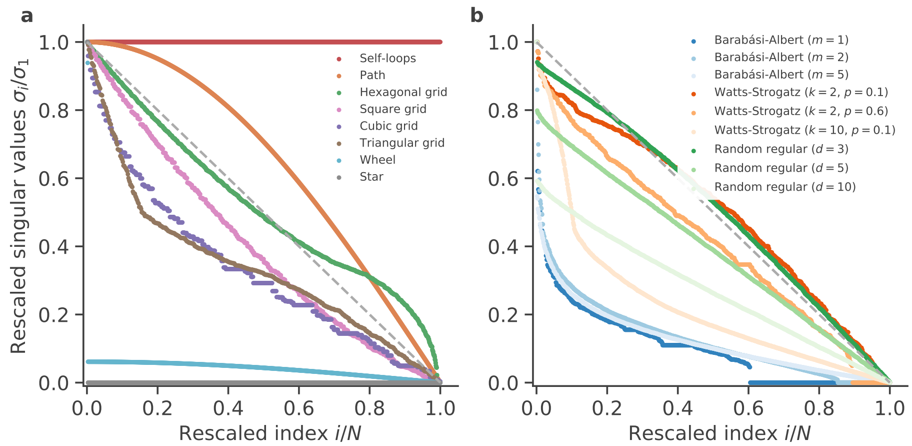

Fig. S2: Rescaled singular values for **a** typical (non-random) graphs and **b** instances of three random graphs with different parameters. For the Barabási-Albert model, m is the number of edges to which a new vertex is attached. For the (connected) Watts-Strogatz model, k is the number of neighbors of each vertex in a ring structure and p is the probability of rewiring an edge. For the random regular model, d is the degree of each vertex. The graphs and random graphs are all available on networkx (see plots/plot_fig_SI_singular_values_scree_graphs.py on the Github repository), except the disconnected self-loop graph whose adjacency matrix is simply the identity matrix. The dashed gray lines are shown for the sake of visualization.

with -1's and 1's for each site that has been rewired with probability p. This is a perfect counter-example of the affirmation "popular random network models are low-rank": the model is a sum of a high-rank matrix and a noise matrix (first indicator). Figure S2b also shows that the singular values in the model can decrease linearly and even supralinearly for some parameters (as in the *random regular graph*). These observations confirm that even though the Watt-Strogatz satisfy two interesting properties, namely small distances (small-world property) and a high-clustering coefficient, it doesn't generally enjoy the low-effective-rank property that we observe in real networks.

Example S18 (Graph theory)**.** Although we often discuss about the rapid decrease of singular values in graphs, they can in fact have very different singular value distributions in general. As an intuitive example, we gather simple graphs, common in physics (e.g., path, grids), and illustrate how their singular values decrease differently, from supralinear to sublinear, in Fig. S2a.

One can also find clear examples of the low-rank models in physics, machine learning, and neuroscience that are worth mentioning.

Example S19 (Random matrix theory and spin glasses)**.** The typical random matrix ensembles used in physics (e.g., *Gaussian Orthogonal Ensemble* [32]) are matrix models involving normally (Gaussian) distributed random variables and such that $\langle W \rangle = 0$, so they have a rank equal to zero. A counter-example is the *Circular Unitary Ensemble* that is of full rank with all the singular values being 1. The random matrix J encoding the interactions in the classical *Sherrington-Kirkpatrick spin-glass model* [117] is formed by i.i.d. Gaussian variables of mean J_0, which implies that $\langle J \rangle = J_0 \mathbf{1}\mathbf{1}^\top$, meaning that the effective rank of the model is one. Other well-known random matrix models, such as the *Gaussian ensembles with finite-rank external source* [118] or *spiked random matrices* [119], satisfy Eq. (S34) with $\langle W \rangle$ of rank $r = O(1)$, i.e., they have a low-rank formulation in the limit $N \to \infty$.

Example S20 (Machine learning and neuroscience)**.**

- In the *Hopfield network* [120], one of the most influential models of artificial recurrent neural network, the weight matrix describing the connections between N dynamical binary units is trained to memorize $n \ll N$ state vectors $\boldsymbol{v}_s \in \{0,1\}^N$. Starting with an initial random symmetric weight matrix of mean zero, T, the training consists in mapping $T \mapsto T + \sum_{s=1}^{n} \boldsymbol{v}_s \boldsymbol{v}_s^\top$, resulting in a final weight matrix of effective rank $\leq n$.
- In an *echo-state network* [121], a random weighted directed graph of mean weight zero is used to generate a reservoir, which is the hidden recurrent part of the artificial neural network that is not affected by learning. The reservoir thus has a rank of zero (according to the first indicator, about the rank of the expected matrix defined in the paper).
- Training *shallow undercomplete autoencoders* is essentially a low-rank approximation problem [22]. The architecture formed by encoding/decoding weight matrices and a hidden layer thus form a low-rank model in itself.
- The *chaotic random neural network* of Sompolinsky et al. [68] is defined using random i.i.d. synaptic weights of mean 0 and variance J^2/N, thus corresponding to an expected weight matrix of rank 0. This model was later

used to set the appropriate initial weights for training RNNs [122]. It has been generalized to include P distinct neuron populations, leading to a rank-P expected weight matrix [103]

- The synaptic weight matrix in the *Rajan-Abbott random neural network* [102] is given by the equation
$$W = \langle W \rangle + J\,,$$
where J is a $N \times N$ random matrix whose elements are i.i.d. of mean 0 and variance $1/N$, while
$$\langle W \rangle = \frac{1}{\sqrt{N}}\mathbf{1}\boldsymbol{v}^\top, \qquad \boldsymbol{v}^\top = (\mu_E, \ldots, \mu_E, \mu_I, \ldots, \mu_I)\,,$$
where f denotes the fraction of excitatory neurons, whose mean synaptic weight is $\mu_E/\sqrt{N} > 0$, while $(1-f)$ denotes the fraction of inhibitory neurons, whose mean synaptic weight is $\mu_I/\sqrt{N} < 0$. The rank of the expected weight matrix is thus equal to 1.
- Another example is the *Gaussian mixture low-rank network* [123–126] whose weight matrix is defined as
$$W = \sum_{r=1}^{R} \boldsymbol{m}_r \boldsymbol{n}_r^\top + R,$$
where R is a zero-mean Gaussian random matrix while $\boldsymbol{m}_r$ and $\boldsymbol{n}_r$ respectively denote the r-th left singular vector of $\langle W\rangle$ and its right singular vector multiplied by the r-th singular value. The rank of $\langle W\rangle$ is thus equal to R. In Ref. [125], for example, the low-rank hypothesis is explicitly made: "We restrict the connectivity matrix to be of low rank, i.e., the number of nonzero singular values of the matrix J is $R \ll N$." In Table I, we call this random graph the "*rank-perturbed Gaussian model*" (RPG) and we absorb the factor $1/N$ in $\boldsymbol{m}_\mu$ and $\boldsymbol{n}_\mu$. Moreover, in Ref.[123], "our theory suggests a simple conjecture: the low-dimensional structure in connectivity determines low-dimensional dynamics and computational properties of recurrent networks." Our paper proves partly this conjecture: by Corollary S74, if $\mathrm{rank}(W) \ll N$ (low-dimensional structure), then the recurrent neural dynamics evolves in a $\mathrm{rank}(W)$-dimensional space (low-dimensional dynamics) (more explanations in Example S76).
- It has also been observed experimentally that trained models have a low effective rank (the ones from NWS [127] in this paper and other references, such as Ref. [128]).

To illustrate the three indicators of the low-rank hypothesis, we use four of the random graphs introduced above, that is RPG, DCSBM, S^1 RGM, and WDSCM, and present the results in Fig. 2 for $N = 10^3$ vertices. Below, we list the parameters used in each model to generate the figure. We denote a N-dimensional realization of a truncated Pareto density $\boldsymbol{x} = p(N, x_{\min}, x_{\max}, \gamma)$, where $N = 10^3$ is the number of vertices, $x_{\min}$ is the minimum value of the distribution, $x_{\max}$ is the maximum value, and γ is the shape parameter. Similarly, an N-dimensional instance of a Gaussian density is denoted $n(N, m, v)$, where m is the mean of the Gaussian and v is its variance, and $u(N, x_{\min}, x_{\max})$ is an N-dimensional instance of discrete uniform distribution.

- For RPG, we set the rank to 5, $\boldsymbol{m}_\mu = n(N, 0, 1/N)$, and $\boldsymbol{n}_\mu = n(N, (\mu-1)/10, \mu/10)$ for all $\mu \in \{1, 2, 3, 4, 5\}$. The $N \times N$ random noise matrix R has elements following a Gaussian density with mean 0 and variance g^2/N where we call g the noise strength. We set $g = 1$ in Fig. 2b and $g = 3$ in Fig. 2f. The parameter g is tuned from 0.01 to 4 in Fig. 2j to increase $\overline{\|R\|}_2$.
- For DCSBM, we set the number of blocks to 5 with respective sizes $N/10$, $2N/5$, $N/10$, $N/5$, and $N/5$. The expected degree distributions are $\boldsymbol{\kappa}_{\mathrm{in}} = p(N, 2, 100, 2.5)$ and $\boldsymbol{\kappa}_{\mathrm{out}} = p(N, 1, 50, 2)$ which are then normalized by groups as defined in Eq. (S62). The expected number of edges is
$$E = gN \begin{pmatrix} 0.40 & 0.10 & 0.10 & 0.02 & 0.13 \\ 0.05 & 0.80 & 0.02 & 0.09 & 0.10 \\ 0.02 & 0.02 & 0.30 & 0.05 & 0.02 \\ 0.10 & 0.05 & 0.05 & 0.40 & 0.01 \\ 0.10 & 0.09 & 0.05 & 0.05 & 0.30 \end{pmatrix},$$
where g is a multiplicative factor that allows increasing $\overline{\|R\|}_2$. We set $g = 100$ in Fig. 2c and $g = 10$ in Fig. 2g. The parameter g is tuned from 200 to 6 to increase $\overline{\|R\|}_2$ in Fig. 2k.
- For S^1 RGM, we set $R = N/2\pi$. To get the angular distance matrix, we set $t_i = 2\pi/[u(N, 1, 50)]_i$ and then compute $\theta_{ij} = \pi - |\pi - |t_i - t_j||$. We observed that the numerical computation of the singular values (and, hence, the rank) of θ is particularly sensitive to the choice of angular matrix and taking a discrete uniform distribution reduced the sensitivity (see tests/test_graphs/test_generate_s1_random_graph and the function "test_thetaij_rank" in the Github repository). We also have $\mu = \beta\sin(\pi/\beta)/(2\pi\langle\boldsymbol{\kappa}\rangle))$ where $\langle\,\rangle$ is the arithmetic mean and $\langle\boldsymbol{\kappa}\rangle = \langle\boldsymbol{\kappa}_{\mathrm{in}}\rangle = \langle\boldsymbol{\kappa}_{\mathrm{out}}\rangle$. The expected degree distributions are $\boldsymbol{\kappa}_{\mathrm{in}} = p(N, 2, 100, 2.5)$, $\boldsymbol{\kappa}_{\mathrm{out}} = p(N, 1, 50, 2)$, and then $\boldsymbol{\kappa}_{\mathrm{out}}$ is redefined to $\boldsymbol{\kappa}_{\mathrm{out}} - \langle\boldsymbol{\kappa}_{\mathrm{out}}\rangle\mathbf{1} + \langle\boldsymbol{\kappa}_{\mathrm{in}}\rangle\mathbf{1}$ $[\langle\boldsymbol{\kappa}_{\mathrm{out}}\rangle < \langle\boldsymbol{\kappa}_{\mathrm{in}}\rangle]$ to ensure that

$\langle \boldsymbol{\kappa}_{\mathrm{in}} \rangle = \langle \boldsymbol{\kappa}_{\mathrm{out}} \rangle$. The temperature $1/\beta$ allows increasing $\overline{\|R\|_2}$ and we tuned it from 0.01 to 0.96 in Fig. 2l. The temperature is 0.2 in Fig. 2d and 0.95 in Fig. 2h.

- For WDSCM, we set $\boldsymbol{y} = p(N, x_{\min}, 0.8, 2.5)$ and $\bar{\boldsymbol{y}} = p(N, x_{\min}, 0.7, 3)$ where $y_{\min} = \bar{y}_{\min} = x_{\min}$ is the parameter that we tune to increase $\overline{\|R\|_2}$. We set $x_{\min} = 0.6$ in Fig. 2e and $x_{\min} = 0.15$ in Fig. 2i. To get Fig. 2m, we increase $x_{\min}$ from 0.1 to 0.65.

In the section "Evidence of the hypothesis for network models" of the paper, we discuss how one can give a more precise perspective for spiked random matrices, such as RPG. Indeed, the singular values of spiked random matrices have a "bulk" related to the singular values of the noise matrix R and the presence of outlying singular values is asymptotically characterized by the Baik-Ben Arous-Péché (BBP) phase transition [129]. Notably, the appearance of $p \leq r$ singular values outliers in W only depends upon a threshold $\bar{\sigma}$ on $\sigma_1(\langle W \rangle), ..., \sigma_r(\langle W \rangle)$ [24]. The simplest case is presented below.

Example S21. Consider that the elements of R are i.i.d. Gaussian white noise of variance $1/N$, then as $N \to \infty$, the singular values of R tend to densely fill the interval $[0, 2]$, the threshold becomes $\bar{\sigma} = 1$, and the i-th singular value of W moves away from the bulk $[0, 2]$ to reach $\sigma_i(\langle W \rangle) + 1/\sigma_i(\langle W \rangle)$ whenever $\sigma_i(\langle W \rangle) > \bar{\sigma}$ for $i \in \{1, ..., r\}$.

Despite the clear dependence of $\langle W \rangle$ over a low-rank matrix L, it is not always clear whether $\langle W \rangle$ has an effective low-rank. This is the case of *soft configuration models* (see Example S15) for which the expected adjacency matrix does not have the explicit form of a rank factorization. In the next section, we introduce the directed soft configuration model as a maximally entropic random graph. Then, we demonstrate that its singular values decrease exponentially rapidly.

B. Exponential decrease of singular values in directed soft configuration models

In general, we only have partial information on complex networks. It is thus reasonable to define a set of networks where each network have a probability to describe the observed complex network. In order to do that in the least biased way, one can rely on the maximization of Shannon entropy to extract an adequate probability distribution [130]. A lot of random graphs are defined from a maximally entropic model and although there is a large literature on the subject [131–134], we provide, for the sake of completeness, some important results and comments. We will later use them to demonstrate Theorem S32 and Theorem S33 on the exponential decrease of singular values in the directed soft configuration model and its weighted version, both maximally entropic random network models.

We begin by presenting general theorems about the use of Lagrange multipliers to obtain maximally entropic network models. Of course, the idea of Lagrange multipliers is old [135, 136]. It goes back to Lagrange and even Euler, but in both of their work, the conditions in which the method applies are not clearly stated and no rigorous demonstration was provided. The first author who clearly stated the theorem is most likely Carathéodory, in the first German edition of his volume on the calculus of variations in 1935 [137, 186 and 187].

Theorem S22 (Lagrange multipliers)**.**
Let:

1. *U, be an open set in $\mathbb{R}^N$;*
2. *$f, g_1, \ldots, g_r$, be continuously differentiable real functions on U;*
3. *E, be a set such that $\boldsymbol{x} \in E$ iff $\boldsymbol{x} \in U$ and $g_1(\boldsymbol{x}) = \cdots = g_r(\boldsymbol{x}) = 0$.*

If $\boldsymbol{x}^*$ maximizes or minimizes f on E, then there exists a real vector $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_r)$ such that:

1. *$\boldsymbol{\lambda} \neq \mathbf{0}$;*
2. *$\lambda_0 \geq 0$;*
3. *$\lambda_0 \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(\boldsymbol{x}^*) = \mathbf{0}$.*

Moreover, if $\nabla g_1(\boldsymbol{x}^*), \ldots, \nabla g_r(\boldsymbol{x}^*)$ are linearly independent, then $\lambda_0 > 0$ and

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{r} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) = \mathbf{0} \tag{S80}$$

for some nonzero vector $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_r^*)$ in $\mathbb{R}^r$.

Proof. The proof is long and often based on the local inversion theorem. See *Carathéodory multiplicative rule* in [138] or [139, Theorem 20.3]. □

Remark S23. On the one hand, the theorem is valid for a minimum or a maximum. This is an advantage that can turn out to be an inconvenience if we do not verify the nature of the point $\boldsymbol{x}^*$. On the other hand, Eq. (S80) is only a *necessary condition* and it is not sufficient in general. We could, for example, find a solution of Eq. (S80) that does not correspond to the desired extremum. Moreover, the theorem supposes that there exists an extremum in E. If this is not assumed, one has to consider an open domain of $\mathbb{R}^N$, which excludes, for example, the compact domain $\bar{D} = [0,1]^N$. Finally, the gradients of the constraints must be linearly independent; otherwise λ_0 can be 0 and that does not help to find the extremum.

The Lagrange multiplier method begins by solving Eq. (S80) by expressing all x_i^* in terms of the multipliers λ_i. Then, the multipliers are written in terms of the known variables by solving the set of constraints $g_1(\boldsymbol{x}) = \cdots = g_n(\boldsymbol{x}) = 0$, which is generally the most difficult step. We finally verify that the solution $\boldsymbol{x}^*$ corresponds to the desired extremum. The following theorem illustrates how the first part of the method can be applied to find the necessary form for the probability mass function P that maximizes the network entropy under (soft) structural constraints.

Theorem S24. *Let A be a $N \times N$ random adjacency matrix with support Ω_A that satisfies the soft equality constraints*

$$\mathbb{E}\left[h_\mu(A)\right] = h_\mu(a^*), \quad \mu \in \{1, ..., \ell\}, \tag{S81}$$

where $\mathbb{E}$ is the expected value on Ω_A, a^* is some $N \times N$ non-random adjacency matrix, and each $h_\mu : \{0,1\}^{N^2} \to \mathbb{R}^\ell$ is continuously differentiable. Then, the probability mass function P that maximizes the entropy of A under the equality constraints (S81) *must be of the form*

$$P(a) = \frac{1}{Z(\lambda_1, ..., \lambda_\ell)} \exp\left[\sum_{\mu=1}^{\ell} \lambda_\mu h_\mu(a)\right], \tag{S82}$$

where $Z : \mathbb{R}^\ell \to \mathbb{R}$ is the partition function and $\lambda_\mu \neq 0$ for all μ.

We can now provide the mathematical steps to show the rapid decrease of singular values in the directed soft configuration model. The next corollary is well known from Ref. [131].

Corollary S25. *Let $\boldsymbol{k}^{\mathrm{in}}$ and $\boldsymbol{k}^{\mathrm{out}}$ be two vectors with elements in $\{1, 2, \ldots, N\}$. Let A be the adjacency matrix of a random directed graph of N vertices, i.e., a random matrix of dimension $N \times N$ and support $\{0,1\}^{N\times N}$. Assume, moreover, that the following constraints are satisfied:*

$$\mathbb{E}[A\mathbf{1}] = \boldsymbol{k}^{\mathrm{in}} \qquad \text{and} \qquad \mathbb{E}[A^\top\mathbf{1}] = \boldsymbol{k}^{\mathrm{out}}\,. \tag{S83}$$

Then the probability mass function P that maximizes the entropy of A must be of the form

$$P(a) = \prod_{i,j=1}^{N} {p_{ij}}^{a_{ij}}(1-p_{ij})^{1-a_{ij}}, \qquad p_{ij} = \frac{\alpha_i\beta_j}{1+\alpha_i\beta_j}, \tag{S84}$$

where α_i and β_j are positive numbers $\forall\, i,j \in \{1, \ldots, N\}$.

Remark S26. The $2N$ scalars $(\alpha_1, ..., \alpha_N, \beta_1, ..., \beta_N)$ are such that $\alpha_i = e^{\lambda_i}$ and $\beta_j = e^{\lambda_{N+j}}$ for all $i, j \in \{1, ..., N\}$ where $\lambda_1, ..., \lambda_N$ are the Lagrange multipliers related to the in-degree constraints, while $\lambda_{N+1}, ..., \lambda_{2N}$ are the Lagrange multipliers related to the out-degree constraints.

Having an explicit form for the probability of a graph in the ensemble allows finding an expression for the expected adjacency matrix, which turns out to have elements following a Fermi-Dirac distribution.

Corollary S27. *Let A be the random matrix described in the previous corollary. Then, for all $i, j \in \{1, \ldots, N\}$,*

$$\langle A_{ij}\rangle = \frac{\alpha_i\beta_j}{1+\alpha_i\beta_j} \tag{S85}$$

and $0 < \langle A_{ij}\rangle < 1$.

The next lemma shows that, under some mild conditions, the expected adjacency matrix is an infinite sum of rank-one matrices with singular values equal to ℓ_1, ℓ_2, $\ldots$ or N, m_1, m_2, m_3, $\ldots$

Lemma S28. *Let A be a random matrix satisfying Eq.* (S85)*. Let*

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^\top, \qquad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)^\top.$$

(1) If $0 < \alpha_i\beta_j < 1$ for all $i, j \in \{1, \ldots, N\}$, then

$$\langle A\rangle = \sum_{k=1}^{\infty} L_k\,, \tag{S86}$$

where L_k denotes a rank-one $N \times N$ matrix whose only nonzero singular value is

$$\ell_k = \sqrt{\sum_{i,j=1}^{N} (\alpha_i \beta_j)^{2k}}\,. \tag{S87}$$

(2) If $\alpha_i\beta_j > 1$ for all $i, j \in \{1, \dots, N\}$, then

$$\langle A \rangle = N\,\hat{\mathbf{1}}\hat{\mathbf{1}}^\top + \sum_{k=1}^{\infty} M_k\,,$$

where M_k denotes a rank-one $N \times N$ matrix whose only nonzero singular value is

$$m_k = \sqrt{\sum_{i,j=1}^{N} (\alpha_i \beta_j)^{-2k}}\,.$$

Proof. This lemma is essentially a direct consequence of expanding the closed form of the geometric series and normalizing the vectors in each term. Indeed, if $0 < \alpha_i\beta_j < 1$ for all $i, j \in \{1, \dots, N\}$, we can use the geometric series and get the following convergent series:

$$\langle A \rangle = \boldsymbol{\alpha}\,\boldsymbol{\beta}^\top - (\boldsymbol{\alpha} \circ \boldsymbol{\alpha})\,(\boldsymbol{\beta} \circ \boldsymbol{\beta})^\top + (\boldsymbol{\alpha} \circ \boldsymbol{\alpha} \circ \boldsymbol{\alpha})\,(\boldsymbol{\beta} \circ \boldsymbol{\beta} \circ \boldsymbol{\beta})^\top\; - + \dots$$

Setting

$$L_k = (-1)^{k+1}(\underbrace{\boldsymbol{\alpha} \circ \cdots \circ \boldsymbol{\alpha}}_{k \text{ times}})(\underbrace{\boldsymbol{\beta} \circ \cdots \circ \boldsymbol{\beta}}_{k \text{ times}})^\top,$$

we get Eq. (S86). We see that each matrix L_k is factorized as $\boldsymbol{u}\boldsymbol{v}^\top$, so we conclude that the rank of each element of the series is one. Moreover, the SVD for such a matrix is simply $\boldsymbol{u}\boldsymbol{v}^\top = \rho\,\widehat{\boldsymbol{u}}\,\widehat{\boldsymbol{v}}^\top$, where $\rho = \|\boldsymbol{u}\|\|\boldsymbol{v}\|$, $\widehat{\boldsymbol{u}} = \boldsymbol{u}/\|\boldsymbol{u}\|$, $\widehat{\boldsymbol{v}} = \boldsymbol{v}/\|\boldsymbol{v}\|$. Hence,

$$L_k = \ell_k\,\widehat{\boldsymbol{\alpha}_k}\,\widehat{\boldsymbol{\beta}_k}^\top, \tag{S88}$$

where

$$\ell_k = \|\underbrace{\boldsymbol{\alpha} \circ \cdots \circ \boldsymbol{\alpha}}_{k \text{ times}}\|\|\underbrace{\boldsymbol{\beta} \circ \cdots \circ \boldsymbol{\beta}}_{k \text{ times}}\|, \qquad \widehat{\boldsymbol{\alpha}_k} = (-1)^{k+1}\frac{\overbrace{\boldsymbol{\alpha} \circ \cdots \circ \boldsymbol{\alpha}}^{k \text{ times}}}{\|\underbrace{\boldsymbol{\alpha} \circ \cdots \circ \boldsymbol{\alpha}}_{k \text{ times}}\|}\,, \qquad \widehat{\boldsymbol{\beta}_k} = \frac{\overbrace{\boldsymbol{\beta} \circ \cdots \circ \boldsymbol{\beta}}^{k \text{ times}}}{\|\underbrace{\boldsymbol{\beta} \circ \cdots \circ \boldsymbol{\beta}}_{k \text{ times}}\|}\,.$$

Simple calculations lead to Eq. (S87), which completes the proof of the first part of the lemma. The second part is proved similarly starting with the geometric series of $\langle A_{ij} \rangle = 1/(1 + 1/(\alpha_i\beta_j))$. □

The last lemma will allow us to find upper bounds on the singular values of the expected adjacency matrix by using Weyl inequalities. However, some technical results are required before deducing the upper bounds. In particular, the coefficients ℓ_k and m_k in Lemma S28 are ordered and bounded as stated in the next lemma.

Lemma S29. *Let ℓ_k and m_k be the coefficients defined in Lemma S28.*
(1) If $0 < \alpha_i\beta_j < 1$ for all $i \in \{1, \dots, N\}$, then

$$\ell_{k+1} < \ell_k\,, \qquad \forall\, k \in \mathbb{Z}_+ \tag{S89}$$

and, with $\gamma = \max_{i,j} \alpha_i\beta_j$,

$$\ell_k \leq N\,\gamma^k\,, \qquad \forall\, k \in \mathbb{Z}_+\,. \tag{S90}$$

(2) If $\alpha_i\beta_j > 1$ for all $i, j \in \{1, \dots, N\}$, then

$$m_{k+1} < m_k\,, \qquad \forall\, k \in \mathbb{Z}_+\,. \tag{S91}$$

and, with $\omega = \min_{i,j} \alpha_i\beta_j$,

$$m_k \leq N\,\omega^{-k}\,, \qquad \forall\, k \in \mathbb{Z}_+\,. \tag{S92}$$

Proof. For the first case, for all $k \in \mathbb{Z}_+$ and from Eq. (S87),

$$\ell_{k+1} = \sqrt{\sum_{i,j=1}^{N} (\alpha_i\beta_j)^{2(k+1)}} = \sqrt{\sum_{i,j=1}^{N} (\alpha_i\beta_j)^{2k}(\alpha_i\beta_j)^2} < \sqrt{\sum_{i,j=1}^{N} (\alpha_i\beta_j)^{2k}} = \ell_k,$$

where we have used $(\alpha_i\beta_j)^2 < 1$ since $0 < \alpha_i\beta_j < 1$ for all $i, j \in \{1, ..., N\}$. The first inequality of case (1) is thus established. Moreover, if $\alpha_i\beta_j \le \gamma$ for all $i, j \in \{1, \dots, N\}$, then

$$\ell_k = \sqrt{\sum_{i,j=1}^{N} (\alpha_i\beta_j)^{2k}} \le \sqrt{\sum_{i,j=1}^{N} \gamma^{2k}} = N\,\gamma^k\,. \tag{S93}$$

The second inequality of case (1) follows froCase (2) is proved similarly. □

Moreover, for a given bound on $\alpha_i\beta_j$, there is a corresponding bound for the elements of the expected adjacency matrix.

Lemma S30. *Let A be a random matrix satisfying Eq. (S85). Let γ and ω be two positive constants. Then,*

$$\alpha_i\beta_j \le \gamma < 1 \iff \langle A_{ij}\rangle \le \frac{\gamma}{1+\gamma} < \frac{1}{2} \qquad \textit{and} \qquad \alpha_i\beta_j \ge \omega > 1 \iff \langle A_{ij}\rangle \ge \frac{\omega}{1+\omega} > \frac{1}{2}. \tag{S94}$$

Proof. Recall that all the parameters involved in this lemma are positive. The first part of both equivalences is obtained with basic inequality manipulations:

$$\alpha_i\beta_j \le \gamma \iff \frac{1}{\alpha_i\beta_j} \ge \frac{1}{\gamma} \iff \frac{1}{1+\frac{1}{\alpha_i\beta_j}} \le \frac{1}{1+\frac{1}{\gamma}} \iff \langle A_{ij}\rangle \le \frac{\gamma}{1+\gamma}\,,$$
$$\alpha_i\beta_j \ge \omega \iff \frac{1}{\alpha_i\beta_j} \le \frac{1}{\omega} \iff \frac{1}{1+\frac{1}{\alpha_i\beta_j}} \ge \frac{1}{1+\frac{1}{\omega}} \iff \langle A_{ij}\rangle \ge \frac{\omega}{1+\omega}\,.$$

The second part is an immediate consequence of $\gamma < 1 \iff \gamma/(1+\gamma) < 1/2$ and $\omega > 1 \iff \omega/(1+\omega) > 1/2$. □

Remark S31. The inequalities in the last lemma imply that for all $i \in \{1, ..., N\}$, the expected degrees k_i^{in} and k_i^{out} are both upper-bounded by $N\gamma/(1+\gamma)$ when $\alpha_i\beta_j < \gamma < 1$, and lower bounded by $N\omega/(1+\omega)$ when $\alpha_i\beta_j > \omega > 1$. However, these bounds on k_i^{in} and k_i^{out} do not necessarily imply that the inequalities in the last lemma are satisfied.

We are now ready to present the first main theorem of this section, which states that for two broad families of parameters defining the soft directed configuration model, the singular values of expected adjacency matrix decrease very rapidly, at least exponentially.

Theorem S32. *Let $\langle A\rangle$ be the $N \times N$ matrix defined in Eq. (S85) and whose singular values are $\sigma_1 \ge ... \ge \sigma_N$. Let ℓ_k and m_k be the coefficients defined in Lemma S28.*

(1) If $0 < \langle A_{ij}\rangle < 1/2$ for all $i, j \in \{1, \dots, N\}$, then the singular values are upper-bounded as

$$\sigma_i \le \sum_{k=i}^{\infty} \ell_k \le \frac{N\,\gamma^i}{1-\gamma}, \qquad \forall\, i \in \{1, ..., N\}\,, \tag{S95}$$

where $\gamma = \max_{i,j} \alpha_i\beta_j$.

(2) If $1/2 < \langle A_{ij}\rangle < 1$ for all $i \in \{1, \dots, N\}$, then the singular values are upper-bounded as

$$\sigma_i \le N\,\delta_{i1} + \sum_{k=i}^{\infty} m_k \le N\delta_{i1} + \frac{N\omega^{1-i}}{\omega - 1}, \qquad \forall\, i \in \{1, ..., N\}, \tag{S96}$$

where δ_{i1} is a Kronecker delta and $\omega = \min_{i,j} \alpha_i\beta_j$.

Proof.
(1) First of all, $0 < \langle A_{ij}\rangle < 1/2$ if and only if $0 < \alpha_i\beta_j < 1$ for all $i, j \in \{1, ..., N\}$ from Lemma S30. Lemma S28 then implies that the expected adjacency matrix is a convergent infinite sum of rank-one matrices L_k, $k \in \mathbb{Z}_+$. Thus, the singular values of the expected adjacency matrix are the singular values of a sum of matrices:

$$\sigma_i(\langle A\rangle) = \sigma_i(\textstyle\sum_{k=1}^{\infty} L_k)\,, \qquad \forall\, i \in \{1, ..., N\},$$

where we write $\sigma_i(\langle A\rangle)$ instead of σ_i for the sake of clarity in the proof.

Next, recall from Theorem S5 that the Weyl inequalities for $N \times N$ matrices B and C are

$$\sigma_{r+s-1}(B + C) \le \sigma_r(B) + \sigma_s(C)\,, \qquad \forall\, 1 \le r,\ s,\ r + s - 1 \le N.$$

Setting $r = s = 1$ yields the familiar triangle inequality:

$$\sigma_1(B + C) \le \sigma_1(B) + \sigma_1(C)\,. \tag{S97}$$

The latter inequality implies that for all $1 \le i \le n-1 < \infty$,

$$\sigma_1\Big(\sum_{k=i}^{n} L_k\Big) \le \sum_{k=i}^{n} \sigma_1(L_k)\,.$$

However, given that $\sigma_1(L_k)$ is nonnegative,

$$\sum_{k=i}^{n} \sigma_1(L_k) \le \sum_{k=i}^{n+1} \sigma_1(L_k) \le \cdots \le \sum_{k=i}^{\infty} \sigma_1(L_k) = \sum_{k=i}^{\infty} \ell_k\,,$$

where we have used the notation $\ell_k = \sigma_1(L_k)$ introduced in Lemma S28. To prove the convergence of the infinite series, we recall from Lemma S29 that $\ell_k \le N\gamma^k$ with $\gamma = \max_{i,j} \alpha_i\beta_j$. This in turn implies that

$$\sum_{k=i}^{\infty} \ell_k \le \frac{N\gamma^i}{1-\gamma},$$

as stated in the rightmost inequality of (S95). So far, we have proved that for all $1 \le i \le n-1 < \infty$,

$$\sigma_1\Big(\sum_{k=i}^{n} L_k\Big) \le \sum_{k=i}^{\infty} \ell_k \le \frac{N\gamma^i}{1-\gamma}\,.$$

The continuity of $\sigma_1 : \mathbb{R}^{N\times N} \to \mathbb{R}$, which is obvious since σ_1 is a norm, and the convergence of $\sum_{k=i}^{\infty} L_k$ allow us to take the limit $n \to \infty$ on the left-hand side of the previous inequality and conclude that

$$\sigma_1\Big(\sum_{k=i}^{\infty} L_k\Big) \le \sum_{k=i}^{\infty} \ell_k \le \frac{N\gamma^i}{1-\gamma}\,. \tag{S98}$$

Let us now go back to the Weyl inequalities and set $r = i$, $s = 1$, $B = \sum_{k=1}^{i-1} L_k$, and $C = \sum_{k=i}^{\infty} L_k$. This yields the inequality

$$\sigma_i\Big(\sum_{k=1}^{\infty} L_k\Big) = \sigma_i\Big(\sum_{k=1}^{i-1} L_k + \sum_{k=i}^{\infty} L_k\Big) \le \sigma_i\Big(\sum_{k=1}^{i-1} L_k\Big) + \sigma_1\Big(\sum_{k=i}^{\infty} L_k\Big)\,,$$

which is valid for all $1 \le i \le N$. The matrix $\sum_{k=1}^{i-1} L_k$ is the sum of $i-1$ matrices of rank one, which means that the rank of $\sum_{k=1}^{i-1} L_k$ is at most $i-1$. Hence, $\sigma_i\Big(\sum_{k=1}^{i-1} L_k\Big) = 0$, so that

$$\sigma_i\Big(\sum_{k=1}^{\infty} L_k\Big) \le \sigma_1\Big(\sum_{k=i}^{\infty} L_k\Big)\,, \qquad \forall\, i \in \{1,\ldots,N\}\,. \tag{S99}$$

Combining inequalities (S98) and (S99) leads to the desired result:

$$\sigma_i\Big(\sum_{k=1}^{\infty} L_k\Big) \le \sum_{k=i}^{\infty} \ell_k \le \frac{N\gamma^i}{1-\gamma}\,, \qquad \forall\, i \in \{1,\ldots,N\}\,.$$

(2) Similarly to the first case Lemmas S28-S30, and Weyl inequalities imply that

$$\sigma_i(\langle A\rangle) = \sigma_i\Big(\sum_{k=1}^{\infty} M_k\Big) \le N\delta_{i1} + m_i + \sigma_1\Big(\sum_{k=i+1}^{\infty} M_k\Big).$$

Proceeding as for inequality (S98) then leads to the inequality

$$\sigma_i\Big(\sum_{k=1}^{\infty} M_k\Big) \le N\delta_{i1} + \sum_{k=i}^{\infty} m_k,$$

where $m_k = \sigma_1(M_k)$. Additionally, Lemma S29 states that $m_k \le N\omega^{-k}$ with $\omega = \min_{i,j} \alpha_i\beta_j$, which leads to

$$\sigma_i\Big(\sum_{k=1}^{\infty} M_k\Big) \le N\delta_{i1} + N\sum_{k=i}^{\infty} \omega^{-k}\,.$$

Writing the truncated geometric series in closed form finally gives the expected result. $\square$

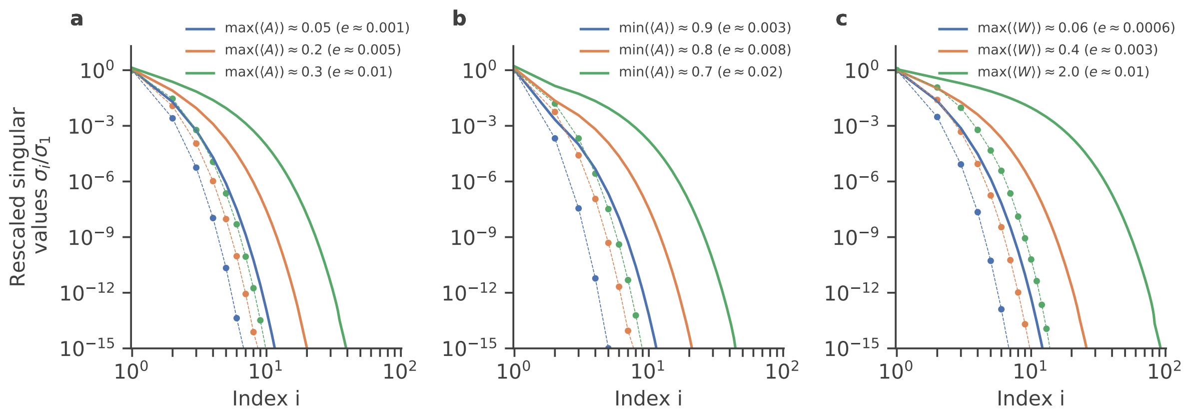

Fig. S3: Upper bounds (solid lines) on the singular values (markers) of the expected matrix of the directed soft configuration model (**a**, Eq. (S95) and **b**, Eq. (S96)) and its weighted version (**c**, Eq. (S100)). The dashed lines between singular values are shown for the sake of visualization. In all the subfigures, the N-dimensional vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{y}$, $\bar{\boldsymbol{y}}$ defining the expected matrices are obtained from truncated Pareto distributions and we denote one N-dimensional realization as $p_{\boldsymbol{x}}(N, x_{\min}, x_{\max}, \gamma)$, where $N = 10^3$ is the number of vertices, $\boldsymbol{x}$ is $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{y}$, or $\bar{\boldsymbol{y}}$, $x_{\min}$ is the minimum value of the distribution, $x_{\max}$ is the maximum value, and γ is the shape parameter. **a**, $p_{\boldsymbol{\alpha}}(N, 2, \alpha_{\max}, 2)/\sqrt{N}$ where $\alpha_{\max} \in \{10, 20, 30\}$ and $p_{\boldsymbol{\beta}}(N, 1, \beta_{\max}, 2.5)/\sqrt{N}$ where $\beta_{\max} \in (5, 10, 15)$. **b**, $p_{\boldsymbol{\alpha}}(N, \alpha_{\min}, 200, 2)/\sqrt{N}$ where $\alpha_{\min} \in (150, 75, 50)$ and $p_{\boldsymbol{\beta}}(N, \beta_{\min}, 150, 2.5)/\sqrt{N}$ where $\beta_{\min} \in (120, 60, 40)$. **c**, $p_{\boldsymbol{y}}(N, 0.05, y_{\max}, 2)$ where $y_{\max} \in \{0.3, 0.6, 0.9\}$ and $p_{\bar{\boldsymbol{y}}}(N, 0.05, \bar{y}_{\max}, 2.5)$ where $\bar{y}_{\max} \in (0.2, 0.5, 0.8)$. All the tuples above correspond to the (blue, orange, green) solid lines.

The upper bounds in the last theorem theoretically validate the low-rank formulation of the directed soft configuration model. The last inequalities in Eqs. (S95) and (S96) are meant to explicitly show the exponential decrease of the singular values while the first inequalities in Eqs. (S95) and (S96) are tighter versions. In Fig. S3a and S3b, we illustrate the first inequalities in Eqs. (S95) and (S96) with both axes in log-log scale.

Following similar steps, we prove the second main theorem of the section: the singular values of $\langle W\rangle$ in the weighted directed soft configuration model (WDSCM) [Example S16] are at least exponentially decreasing.

Theorem S33. *Let $\langle W\rangle$ be the $N\times N$ matrix defined in Eq. (S72) and whose singular values are $\sigma_1 \geq ... \geq \sigma_N$ and let $n_k = \sqrt{\sum_{i,j=1}^N (y_i\bar{y}_j)^{2k}}$ with $0 < y_i\bar{y}_j < 1$ for all i, j. Then, the singular values are upper-bounded as*

$$\sigma_i \leq \sum_{k=i}^{\infty} n_k \leq \frac{N\,\tau^i}{1-\tau}, \qquad \forall\, i \in \{1, ..., N\}\,, \tag{S100}$$

where $\tau = \max_{i,j} y_i\bar{y}_j$.

Contrarily to Theorem S32, there is no restriction on the domain of the elements of $\langle W\rangle$ for the inequality (S100), which is a consequence of the Bose-Einstein distribution for the elements of the expected weight matrix. The bound in Eq. (S100) is illustrated in Fig. S3c.

C. Impact of singular value distribution and matrix density on effective ranks

In this subsection, we take advantage of the formula for the stable rank (srank), the nuclear rank (nrank), and the erank, which are amenable for analytic calculations, to assess the impact of different singular value decreases on the effective ranks through various inequalities. In the first part of the subsection, we prove that finding bounding curves, $\psi_*(x)$ and $\psi^*(x)$, that approximately delineate the region of possible singular value allows us to estimate the srank, nrank, and erank. In the second part, we show that linear $O(N)$, sublinear $O(N^{1-\epsilon})$, and constant $O(1)$ asymptotic behaviors emerge depending on the shape of the singular value distribution. We finally present, in a third part, the impact of the density of W on srank(W) through general inequalities.

1. Singular-value envelopes

We define the *singular-value envelopes* ψ_* and ψ^* for the singular values as functions that

1. decrease on the interval $[1, N] \subset \mathbb{R}$, that is

$$x \leq y \qquad \Longrightarrow \qquad \psi_*(x) \geq \psi_*(y), \qquad \psi^*(x) \geq \psi^*(y)\,; \tag{S101}$$

2. are nonnegative on the interval $[1, N] \subset \mathbb{R}$;
3. provide lower and upper bounds for the rescaled singular values as

$$\psi_*(i) \le \frac{\sigma_i}{\sigma_1} \le \psi^*(i) \qquad \forall\, i \in \{1, \dots, N\}; \tag{S102}$$

4. are σ_1–tight, meaning

$$\psi_*(1) = 1 = \psi^*(1)\,. \tag{S103}$$

The last condition is imposed to always match the only value of the ratio σ_i/σ_1 that is known in all instances. In the next part of the subsection, it will also prevent us from multiplying ψ_* and ψ^* by global scaling factors such as $N^{-\epsilon}$, which could impose, somewhat artificially, asymptotic behaviors for the effective ranks such as $O(N^{1-\epsilon})$. The four properties of the singular-value envelopes readily imply general inequalities that will be useful to bound the effective ranks.

Lemma S34 (Basic inequalities)**.** *If ψ_* and ψ^* satisfy conditions 1–4, then*

$$\int_1^N \psi_*(x)^q\, dx \;\le\; \sum_{i=1}^N \psi_*(i)^q \;\le\; \sum_{i=1}^N \left(\frac{\sigma_i}{\sigma_1}\right)^q \;\le\; \sum_{i=1}^N \psi^*(i)^q \;\le\; 1 + \int_1^N \psi^*(x)^q\, dx \tag{S104}$$

for all $q \ge 0$ and

$$0 \;\le\; \sum_{i=2}^N \psi_*(i) \ln \frac{1}{\psi^*(i)} \;\le\; \sum_{i=1}^N \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \;\le\; \frac{1}{\delta} \sum_{i=2}^N \big(\psi^*(i)^{1-\delta} - \psi_*(i)\big) \;\le\; \frac{1}{\delta} \int_1^N \big(\psi^*(x)^{1-\delta} - \psi_*(x)\big)\, dx + \Delta \tag{S105}$$

for all $\delta \in (0, 1)$, where

$$\Delta = \frac{1}{\delta} \int_1^2 \psi_*(x)\, dx\,. \tag{S106}$$

Proof. We first prove the inequalities involving the summations. Using inequality (S102) and the nonnegativeness of ψ_*, ψ^*, and σ_i/σ_1, we deduce the inequality

$$\psi_*(i)^q \;\le\; \left(\frac{\sigma_i}{\sigma_1}\right)^q \;\le\; \psi^*(i)^q, \tag{S107}$$

which is valid for all $q \ge 0$ and $i \in \{1, 2, \dots, N\}$. Thus

$$\sum_{i=1}^N \psi_*(i)^q \;\le\; \sum_{i=1}^N \left(\frac{\sigma_i}{\sigma_1}\right)^q \;\le\; \sum_{i=1}^N \psi^*(i)^q\,, \tag{S108}$$

as expected.

We now concentrate on the first and last inequalities. We adopt a strategy analogous to the method used for proving the integral test for convergence. Notice that the ψ_* and ψ^* are integrable on any subinterval of $[1, N]$ since these functions are monotone. On the one hand,

$$\psi_*(i)^q = \int_i^{i+1} \psi_*(i)^q\, dx \ge \int_i^{i+1} \psi_*(x)^q\, dx \tag{S109}$$

since $\psi_*(i)^q \ge \psi_*(x)^q$ for all $x \ge i$ (condition 1). The latter inequality and condition 2 then imply

$$\begin{aligned} \sum_{i=1}^N \psi_*(i)^q &= \sum_{i=1}^{N-1} \psi_*(i)^q + \psi_*(N)^q \\ &\ge \sum_{i=1}^{N-1} \int_i^{i+1} \psi_*(x)^q\, dx + \psi_*(N)^q = \int_1^N \psi_*(x)^q\, dx + \psi_*(N)^q \ge \int_1^N \psi_*(x)^q\, dx\,, \end{aligned} \tag{S110}$$

which proves the leftmost inequality. On the other hand,

$$\psi^*(i)^q = \int_{i-1}^i \psi^*(i)^q\, dx \le \int_{i-1}^i \psi^*(x)^q\, dx \tag{S111}$$

since $\psi^*(i)^q \leq \psi^*(x)^q$ for all $x \leq i$ (condition 1). Thus, with condition 4,

$$\sum_{i=1}^{N} \psi^*(i)^q = \psi^*(1)^q + \sum_{i=2}^{N} \psi^*(i)^q \leq \psi^*(1)^q + \sum_{i=2}^{N} \int_{i-1}^{i} \psi^*(x)^q \, dx = 1 + \int_{1}^{N} \psi^*(x)^q \, dx \,, \tag{S112}$$

which establishes the rightmost inequality.

To prove the last inequalities, we first notice that thanks to condition 4, the term corresponding to $i = 1$ in the summation can be omitted:

$$\sum_{i=1}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} = \sum_{i=2}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i}. \tag{S113}$$

We then lower-bound each element of the sum as

$$\frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \geq 0 \tag{S114}$$

since $\sigma_1/\sigma_i \geq 1$ for all i. To get the upper bound of (S113), we use the classical inequality $\ln x \leq a(x^{1/a} - 1)$ for all $a, x > 0$ [140, (4.5.5)], which implies that

$$\frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \leq \frac{1}{\delta} \frac{\sigma_i}{\sigma_1} \left(\frac{\sigma_1}{\sigma_i} \right)^{\delta} - \frac{1}{\delta} \frac{\sigma_i}{\sigma_1} = \frac{1}{\delta} \left(\frac{\sigma_i}{\sigma_1} \right)^{1-\delta} - \frac{1}{\delta} \frac{\sigma_i}{\sigma_1} \qquad \forall\, \delta > 0, \tag{S115}$$

where the equality is obtained when $\delta \to 0$. Thus,

$$\begin{aligned} \sum_{i=2}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} &\leq \frac{1}{\delta} \sum_{i=2}^{N} \left(\frac{\sigma_i}{\sigma_1} \right)^{1-\delta} - \frac{1}{\delta} \sum_{i=2}^{N} \frac{\sigma_i}{\sigma_1} \\ &\leq \frac{1}{\delta} \sum_{i=2}^{N} \psi^*(i)^{1-\delta} - \frac{1}{\delta} \sum_{i=2}^{N} \psi_*(i) \leq \frac{1}{\delta} \sum_{i=2}^{N} \psi^*(i)^{1-\delta} - \frac{1}{\delta} \sum_{i=2}^{N-1} \psi_*(i) \end{aligned} \tag{S116}$$

where the second inequality is the expected result while the third one is obtained by neglecting the last (negative) element of the sum. Now, considering that $\psi^*(x)^{1-\delta}$ (with $\delta < 1$) and $-\psi_*(x)$ are respectively decreasing and increasing in x, we can write

$$\sum_{i=2}^{N} \psi^*(i)^{1-\delta} = \sum_{i=2}^{N} \int_{i-1}^{i} \psi^*(i)^{1-\delta} dx \leq \sum_{i=2}^{N} \int_{i-1}^{i} \psi^*(x)^{1-\delta} dx = \int_{1}^{N} \psi^*(x)^{1-\delta} dx, \qquad \forall\, \delta \in (0,1) \tag{S117}$$

and

$$-\sum_{i=2}^{N-1} \psi_*(i) = -\sum_{i=2}^{N-1} \int_{i}^{i+1} \psi_*(i) \, dx \leq -\sum_{i=2}^{N-1} \int_{i}^{i+1} \psi_*(x) \, dx = -\int_{2}^{N} \psi_*(x) \, dx \tag{S118}$$

Hence,

$$\sum_{i=2}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \leq \frac{1}{\delta} \left(\int_{1}^{N} \psi^*(x)^{1-\delta} dx - \int_{2}^{N} \psi_*(x) \, dx \right), \tag{S119}$$

which is equivalent to the desired result. □

In Theorems S32 and S33, we observed exponential decreases of the singular values occurs when working with the expected adjacency or weight matrix of two frequently used random graphs, namely the directed soft configuration model (DSCM) and its weighted version (WDSCM). A first consequence of the previous lemma is that an exponential decrease implies that srank, nrank, and erank are bounded by finite geometric series (or functions of them).

Proposition S35 (Bounds on effective ranks – Exponential decrease)**.** *Suppose that the singular values of matrix W, $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_N \geq 0$, satisfy the inequalities*

$$\alpha^{i-1} \leq \frac{\sigma_i}{\sigma_1} \leq \omega^{i-1}, \qquad i \in \{1, \ldots, N\}, \tag{S120}$$

for some $0 < \alpha \le \omega < 1$. *Then,*

$$g(\alpha^2, N) \le \mathrm{srank}(W) \le g(\omega^2, N) \tag{S121}$$

$$g(\alpha, N) \le \mathrm{nrank}(W) \le g(\omega, N), \tag{S122}$$

$$\omega\, g(\alpha, N)\, \exp\left(\frac{\alpha\, g'(\alpha, N)}{g(\omega, N)}\right) \le \mathrm{erank}(W) \le \alpha\, g(\omega, N)\, \exp\left(\frac{\omega\, g'(\omega, N)}{g(\alpha, N)}\right), \tag{S123}$$

where

$$g(\alpha, N) = \frac{1-\alpha^N}{1-\alpha}, \qquad g'(\alpha, N) = \frac{\partial g(\alpha, N)}{\partial \alpha} = \frac{1+(N-1)\alpha^N - \alpha^{N+1}}{(1-\alpha)^2}. \tag{S124}$$

Proof. The inequalities for nrank and srank are easily derived from Lemma S34 by setting $q = 1$ and $q = 2$, respectively, together with $\psi_*(i) = \alpha^{i-1}$ and $\psi^*(i) = \omega^{i-1}$. Then, combining the first inequalities with Eq. (S33), we obtain the following preliminary result:

$$g(\alpha, N) \exp\left(\frac{1}{g(\omega, N)} \sum_{i=1}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i}\right) \le \mathrm{erank}(W) \le g(\omega, N) \exp\left(\frac{1}{g(\alpha, N)} \sum_{i=1}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i}\right), \tag{S125}$$

where it is understood that $0 \ln 0 = 0$. Now,

$$\alpha^{i-1} \ln \omega^{i-1} \le \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \le \omega^{i-1} \ln \alpha^{i-1}$$

for all $i \in \{1, \ldots, N\}$. Hence,

$$\sum_{i=1}^{N} \alpha^{i-1} \ln \omega^{i-1} \le \sum_{i=1}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \le \sum_{i=1}^{N} \omega^{i-1} \ln \alpha^{i-1},$$

which can be simplified as

$$\alpha \ln \omega \sum_{j=0}^{N-1} j\, \alpha^{j-1} \le \sum_{i=1}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \le \omega \ln \alpha \sum_{j=0}^{N-1} j\, \omega^{j-1}.$$

Using the geometric series and its derivative, we deduce that

$$\alpha \ln \omega\, g'(\alpha, N) \le \sum_{i=1}^{N} \frac{\sigma_i}{\sigma_1} \ln \frac{\sigma_1}{\sigma_i} \le \omega \ln \alpha\, g'(\omega, N).$$

We finally get the desired result by taking the exponential of the previous expression and returning to inequality (S125). □

In Lemma S34, the variable x interpolates between the singular value indices i and thus belongs to a domain that grows with N, which is particularly convenient for studying random network models individually, such as the soft configuration model. To allow the comparison of real networks of different sizes, as in Fig. 1e, we also need to treat the case where x belongs to the closed interval $[0, 1]$, a compact domain that remains the same for all N. The following lemma allows one to go from one perspective to the other.

Lemma S36 (Extensive vs. intensive domains)**.** *Let* $\psi : [1, N] \to [0, 1]$ *be monotonically decreasing. Define* $\Psi : [0, 1] \to [0, 1]$ *as*

$$\Psi(y) = \psi\big((N-1)y + 1\big) \qquad \forall\, y \in [0, 1]. \tag{S126}$$

Then, for all $q > 0$,

$$\int_1^N \psi(x)^q\, dx = (N-1) \int_0^1 \Psi(y)^q\, dy, \qquad \int_1^2 \psi(x)^q\, dx = (N-1) \int_0^{\frac{1}{N-1}} \Psi(y)^q\, dy, \tag{S127}$$

Proof. The first result is an immediate consequence of the following linear, and thus invertible, change of variables:

$$T : [1, N] \longrightarrow [0, 1] \tag{S128}$$

$$x \longmapsto y = \frac{x-1}{N-1} \tag{S129}$$

which maps $[\ell, m]$ onto $[(\ell-1)/(N-1),\ (m-1)/(N-1)]$ for all $1 \le \ell \le m \le N$. □

The two previous lemmas and formula (S33) imply the following result stating that srank and nrank are essentially equal to N times the area under the curves $\Psi(y)^2$ and $\Psi(y)$ with $y \in [0,1]$, respectively, while erank is related to the area under the curves $\Psi(y)$ and $\ln \Psi(y)^{-1}$.

Lemma S37 (Effective rank as area under a curve)**.** *Let ψ_* and ψ^* satisfy conditions 1–4. Define $\Psi_* : [0,1] \to [0,1]$ and $\Psi^* : [0,1] \to [0,1]$ as*

$$\Psi_*(y) = \psi_*\big((N-1)y+1\big) \qquad \text{and} \qquad \Psi^*(y) = \psi^*\big((N-1)y+1\big). \tag{S130}$$

Then

$$(N-1)\int_0^1 \Psi_*(y)^2\, dy \leq \operatorname{srank}(W) \ \leq\ 1 + (N-1)\int_0^1 \Psi^*(y)^2\, dy\,, \tag{S131}$$

$$(N-1)\int_0^1 \Psi_*(y)\, dy \leq \operatorname{nrank}(W) \ \leq 1 + (N-1)\int_0^1 \Psi^*(y)\, dy\,. \tag{S132}$$

Moreover, for all $\delta \in (0,1)$,

$$(N-1)\int_0^1 \Psi_*(y)\, dy \ \leq\ \operatorname{erank}(W) \ \leq\ \left(1 + (N-1)\int_0^1 \Psi^*(y)\, dy\right) e^{\mathcal{H}+\mathcal{P}}\,, \tag{S133}$$

where

$$\mathcal{H} = \frac{1}{\delta}\left(\frac{\displaystyle\int_0^1 \Psi^*(y)^{1-\delta}\, dy}{\displaystyle\int_0^1 \Psi_*(y)\, dy} - 1\right), \qquad \mathcal{P} = \frac{\displaystyle\int_0^{\frac{1}{N-1}} \Psi_*(y)\, dy}{\displaystyle\delta \int_0^1 \Psi_*(y)\, dy}\,. \tag{S134}$$

This new perspective on the effective ranks allows us to consider a general family of singular-value envelopes that can be applied to our experimental results as illustrated in Fig. 1e. Interestingly, this family is related to the Gaussian hypergeometric function [140, Chap. 15].

Theorem S38 (Bounds on effective ranks – Hypergeometric decrease)**.** *Suppose that the singular values of matrix W, $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_N \geq 0$, satisfy the inequality*

$$\left(1 - \frac{i-1}{N-1}\right)^{c^*-2}\left(1 + \zeta^* \frac{i-1}{N-1}\right)^{-b^*} \ \leq\ \frac{\sigma_i}{\sigma_1} \ \leq\ \left(1 - \frac{i-1}{N-1}\right)^{c_*-2}\left(1 + \zeta_* \frac{i-1}{N-1}\right)^{-b_*} \tag{S135}$$

for some $0 \leq b_* \leq b^*$, $2 \leq c_* \leq c^*$, $0 < \zeta_* \leq \zeta^*$, and for all $i \in \{1, \ldots, N\}$. Then,

$$\frac{N-1}{2c^*-3}\ {}_2F_1(1, 2b^*; 2c^*-2; -\zeta^*) \ \leq \operatorname{srank}(W) \ \leq 1 + \frac{N-1}{2c_*-3}\ {}_2F_1(1, 2b_*; 2c_*-2; -\zeta_*) \tag{S136}$$

$$\frac{N-1}{c^*-1}\ {}_2F_1(1, b^*; c^*; -\zeta^*) \leq \operatorname{nrank}(W) \ \leq 1 + \frac{N-1}{c_*-1}\ {}_2F_1(1, b_*; c_*; -\zeta_*), \tag{S137}$$

$$\frac{N-1}{c^*-1}\ {}_2F_1(1, b^*; c^*; -\zeta^*) \leq \operatorname{erank}(W) \ \leq \left(1 + \frac{N-1}{c_*-1}\ {}_2F_1(1, b_*; c_*; -\zeta_*)\right) e^{\mathcal{H}+\mathcal{P}}, \tag{S138}$$

where, for all $\delta \in (0,1)$,

$$\mathcal{H} = \frac{1}{\delta}\left(\frac{c^*-1}{(1-\delta)c_* + 2\delta - 1}\ \frac{{}_2F_1(1, (1-\delta)b_*; (1-\delta)c_* + 2\delta; -\zeta_*)}{{}_2F_1(1, b^*; c^*; -\zeta^*)} - 1\right), \tag{S139}$$

$$\mathcal{P} = \frac{1}{N-1}\frac{(c^*-1)\rho(b^*, c^*, \zeta^*)}{\delta\ {}_2F_1(1, b_*; c_*; -\zeta_*)}\,, \tag{S140}$$

with $\rho(b^*, c^*, \zeta^*)$ being bounded as $0 \leq \rho(b^*, c^*, \zeta^*) \leq 1$ for all $b^* \geq 0$ and $c^* \geq 2$.

Proof. We apply Lemma S37 to the case where the enveloping functions ψ_* and ψ^* are defined as

$$\psi_*(x) = \left(1 - \frac{x-1}{N-1}\right)^{c^*-2}\left(1 + \zeta^* \frac{x-1}{N-1}\right)^{-b^*}, \qquad \psi^*(x) = \left(1 - \frac{x-1}{N-1}\right)^{c_*-2}\left(1 + \zeta_* \frac{x-1}{N-1}\right)^{-b_*}.$$

Changing the variable x for $y = (x-1)/(N-1)$, we get the functions

$$\Psi_*(y) = \frac{(1-y)^{c^*-2}}{(1+\zeta^* y)^{b^*}}, \qquad \Psi^*(y) = \frac{(1-y)^{c_*-2}}{(1+\zeta_* y)^{b_*}}\,.$$

Now, using the integral representation of the hypergeometric function ${}_2F_1(a,b;c;z)$ [140, (15.6.1)] and the symmetry property ${}_2F_1(a,b;c;z) = {}_2F_1(b,a;c;z)$, we get the following formulas:

$$\int_0^1 \frac{(1-y)^{c-2}}{(1+\zeta y)^b}\,dy = \frac{1}{c-1}\,{}_2F_1(1,b;c;-\zeta) \qquad \int_0^1 \left(\frac{(1-y)^{c-2}}{(1+\zeta y)^b}\right)^2 dy = \frac{1}{2c-3}\,{}_2F_1(1,2b;2c-2;-\zeta)\,.$$

Thus, according to the bounds for nrank and srank provided in Lemma S37,

$$\frac{N-1}{c^*-1}\,{}_2F_1(1,b^*;c^*;-\zeta^*) \le \mathrm{nrank}(W) \le 1 + \frac{N-1}{c_*-1}\,{}_2F_1(1,b_*;c_*;-\zeta_*)\,,$$
$$\frac{N-1}{2c^*-3}\,{}_2F_1(1,2b^*;2c^*-2;-\zeta^*) \le \mathrm{srank}(W) \le 1 + \frac{N-1}{2c_*-3}\,{}_2F_1(1,2b_*;2c_*-2;-\zeta_*)\,,$$

as expected. Moreover, according to the lower bound in inequality (S133),

$$\mathrm{erank}(W) \ge \frac{N-1}{c^*-1}\,{}_2F_1(1,b^*;c^*;-\zeta^*)$$

To get the upper bound of erank(W), we use once again inequality (S133):

$$\mathrm{erank}(W) \le \left(1 + \frac{N-1}{c_*-1}\,{}_2F_1(1,b_*;c_*;-\zeta_*)\right) e^{\mathcal{H}+\mathcal{P}}.$$

where $\mathcal{H}$ and $\mathcal{P}$ remain to be evaluated from Eq. (S134). One the one hand, the integral

$$\int_0^1 \left(\frac{(1-y)^{c_*-2}}{(1+\zeta_* y)^{b_*}}\right)^{1-\delta} dy = \frac{1}{(1-\delta)c_*+2\delta-1}\,{}_2F_1(1,(1-\delta)b_*;(1-\delta)c_*+2\delta;-\zeta_*)$$

yields Eq. (S139). On the other hand, the bounding inequality $\Psi_*(y) \le 1$, valid for all $b^* \ge 1$ and $c^* \ge 2$ implies that

$$\int_0^{\frac{1}{N-1}} \Psi_*(y)\,dy \le \int_0^{\frac{1}{N-1}} dy = \frac{1}{N-1}\,.$$

This allows us to define the function ρ such that

$$\rho(b^*,c^*,\zeta^*) = (N-1)\int_0^{\frac{1}{N-1}} \Psi_*(y)\,dy\,, \qquad 0 \le \rho(b^*,c^*,\zeta^*) \le 1\,,$$

and Eq. (S140) follows along with the theorem. □

The singular-value envelopes in the latter theorem are general in the sense that they include, as particular cases, sub-linear, linear, supra-linear, and power-law decreases or mixes of those shapes. As auxiliary result, we provide the following proposition for the the sub- to supra-linear decreases which is a direct implication of Lemma S37.

Proposition S39 (Bounds on effective ranks for sub-linear to supra-linear decrease)**.** *Suppose that the singular values of matrix W, $\sigma_1 \ge \sigma_2 \ge \cdots \sigma_N \ge 0$, satisfy the inequalities*

$$\left(1 - a\frac{i-1}{N-1}\right)^b \le \frac{\sigma_i}{\sigma_1} \le \left(1 - c\frac{i-1}{N-1}\right)^d, \qquad i \in \{1,\dots,N\}, \tag{S141}$$

for some $0 < c \le a \le 1$ and $0 < d \le b$. Then,

$$(N-1)\,\ell(a,2b,1) \le \mathrm{srank}(W) \le 1 + (N-1)\,\ell(c,2d,1)\,, \tag{S142}$$
$$(N-1)\,\ell(a,b,1) \le \mathrm{nrank}(W) \le 1 + (N-1)\,\ell(c,d,1)\,, \tag{S143}$$
$$(N-1)\,\ell(a,b,1) \le \mathrm{erank}(W) \le [1 + (N-1)\,\ell(c,d,1)]\,e^{\mathcal{H}+\mathcal{P}}\,, \tag{S144}$$

where, for all $\delta \in (0,1)$,

$$\ell(a,b,\alpha) = \frac{1-(1-\alpha\,a)^{1+b}}{a(1+b)}, \qquad \mathcal{H} = \frac{1}{\delta}\left(\frac{\ell(c,d(1-\delta),1)}{\ell(a,b,1)} - 1\right), \qquad \textit{and} \qquad \mathcal{P} = \frac{\ell(a,b,\frac{1}{N-1})}{\delta\,\ell(a,b,1)}.$$

The results in this part of the subsection only depend on the curves enveloping the singular values and can thus be used for observed singular values of real networks or to random matrix/graph models. In the following, we relate each singular value decreases to asymptotic behaviors in random graphs.

2. Asymptotic behaviors of the effective ranks in growing graphs

We start this part by highlighting a striking consequence of Proposition S35 : if the singular values decrease exponentially, then the basic effective ranks are $O(1)$ as $N \to \infty$. Thus, the effective rank to dimension ratios are negligible as $N \to \infty$. This is precisely stated in the next corollary.

Corollary S40 (Exponential decrease implies $O(1)$ effective ranks)**.** *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 $0 < \alpha \le \omega < 1$ and for each N, the singular values $\sigma_1 \ge \sigma_2 \ge \cdots \sigma_N \ge 0$ of W_N satisfy the inequalities*

$$\alpha^{i-1} \le \frac{\sigma_i}{\sigma_1} \le \omega^{i-1}, \qquad i \in \{1, \dots, N\}. \tag{S145}$$

Then, as $N \to \infty$,

$$\frac{1}{1-\alpha^2} + O(\alpha^{2N}) \le \operatorname{srank}(W_N) \le \frac{1}{1-\omega^2} + O(\omega^{2N}), \tag{S146}$$

$$\frac{1}{1-\alpha} + O(\alpha^{N}) \le \operatorname{nrank}(W_N) \le \frac{1}{1-\omega} + O(\omega^{N}), \tag{S147}$$

$$\frac{\omega}{1-\alpha}\exp\left(\frac{\alpha(1-\omega)}{(1-\alpha)^2}\right) + O\left(\max\{\omega^N, N\alpha^N\}\right) \le \operatorname{erank}(W_N) \le \frac{\alpha}{1-\omega}\exp\left(\frac{\omega(1-\alpha)}{(1-\omega)^2}\right) + O(N\omega^N). \tag{S148}$$

Proof. We essentially expand the bounds of Proposition S35 and look for the first subdominant terms as $N \to \infty$. On the one hand,

$$g(\alpha^k, N) = \frac{1-\alpha^{kN}}{1-\alpha^k} = \frac{1}{1-\alpha^k}(1-\alpha^{kN}).$$

Hence,

$$\lim_{N\to\infty} \frac{\left| g(\alpha^k, N) - \dfrac{1}{1-\alpha^k} \right|}{\alpha^{kN}} = \frac{1}{1-\alpha^k} < \infty$$

meaning that

$$g(\alpha^k, N) = \frac{1}{1-\alpha^k} + O(\alpha^{kN}).$$

The last asymptotic development readily implies the bounds for srank and nrank. On the other hand,

$$\frac{\alpha\, g'(\alpha, N)}{g(\omega, N)} = \frac{\alpha(1-\omega)}{(1-\alpha)^2}\left(1 + N\alpha^N - \alpha^N - \alpha^{N+1}\right)\left(1-\omega^N\right)^{-1},$$

which allows computing the limit

$$\lim_{N\to\infty} \frac{\left| \dfrac{\alpha\, g'(\alpha, N)}{g(\omega, N)} - \dfrac{\alpha(1-\omega)}{(1-\alpha)^2\,(1-\omega^N)} \right|}{N\alpha^N} = \frac{\alpha(1-\omega)}{(1-\alpha)^2} < \infty.$$

Hence,

$$\frac{\alpha\, g'(\alpha, N)}{g(\omega, N)} = \frac{\alpha(1-\omega)}{(1-\alpha)^2\,(1-\omega^N)} + O\left(N\alpha^N\right).$$

However,

$$\frac{\alpha(1-\omega)}{(1-\alpha)^2\,(1-\omega^N)} = \frac{\alpha(1-\omega)}{(1-\alpha)^2} + O(\omega^N)$$

since

$$\lim_{N\to\infty} \frac{\left| \dfrac{\alpha(1-\omega)}{(1-\alpha)^2\,(1-\omega^N)} - \dfrac{\alpha(1-\omega)}{(1-\alpha)^2} \right|}{\omega^N} = \frac{\alpha(1-\omega)}{(1-\alpha)^2} \lim_{N\to\infty} \frac{\left| \dfrac{\omega^N}{1-\omega^N} \right|}{\omega^N} = \frac{\alpha(1-\omega)}{(1-\alpha)^2}$$

Thus,

$$\frac{\alpha\, g'(\alpha, N)}{g(\omega, N)} = \frac{\alpha(1-\omega)}{(1-\alpha)^2} + O(\omega^N) + O\left(N\alpha^N\right) = \frac{\alpha(1-\omega)}{(1-\alpha)^2} + O\left(\max\{\omega^N, N\alpha^N\}\right),$$

where we have invoked the basic property $O(u) + O(v) = O(\max\{u, v\})$. Consequently, the lower bound of erank has the following asymptotic expansion:

$$\begin{aligned}\omega\, g(\alpha,N)\,\exp\left(\frac{\alpha\, g'(\alpha,N)}{g(\omega,N)}\right) &= \left(\frac{\omega}{1-\alpha}+O(\alpha^N)\right)\exp\left(\frac{\alpha(1-\omega)}{(1-\alpha)^2}+O\left(\max\{\omega^N, N\alpha^N\}\right)\right)\\ &= \left(\frac{\omega}{1-\alpha}+O(\alpha^N)\right)\exp\left(\frac{\alpha(1-\omega)}{(1-\alpha)^2}\right)\left(1+O\left(\max\{\omega^N, N\alpha^N\}\right)\right)\\ &= \left(\frac{\omega}{1-\alpha}+O\left(\max\{\omega^N, N\alpha^N\}\right)\right)\exp\left(\frac{\alpha(1-\omega)}{(1-\alpha)^2}\right),\end{aligned}$$

where the second line has been deduced using the well-known asymptotic formulas $O(u)O(v) = O(uv)$ and $O(u) + O(v) = O(\max\{u, v\})$. The upper bound for erank is obtained from the last result by permuting α and ω, and considering $O(\max\{N\omega^N, \alpha^N\}) = O(N\omega^N)$. $\square$

In light of Lemma S37, which bounds the effective ranks with terms proportional to $N-1$, the previous asymptotic behavior was rather surprising. On the contrary, the next result is fully expected: slowly decreasing envelopes lead to effective ranks that grow linearly with N.

Corollary S41 (Sub-linear to supra-linear decrease imply $O(N)$ effective ranks)**.** *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 a, b, c, d such that $0 < c \le a \le 1$ and $0 < d \le b$, and for each N, the singular values $\sigma_1 \ge \sigma_2 \ge \cdots \sigma_N \ge 0$ of W_N satisfy the inequalities*

$$\left(1-a\,\frac{i-1}{N-1}\right)^b \le \frac{\sigma_i}{\sigma_1} \le \left(1-c\,\frac{i-1}{N-1}\right)^d, \qquad i\in\{1,\ldots,N\}\,. \tag{S149}$$

Then, as $N\to\infty$ and for all $\delta\in(0,1)$,

$$N\,\ell(a,2b,1)+O(1) \le \operatorname{srank}(W_N) \le N\,\ell(c,2d,1)+O(1)\,, \tag{S150}$$

$$N\,\ell(a,b,1)+O(1) \le \operatorname{nrank}(W_N) \le N\,\ell(c,d,1)+O(1), \tag{S151}$$

$$N\,\ell(a,b,1)+O(1) \le \operatorname{erank}(W_N) \le N\,\ell(c,d,1)\exp\left[\frac{1}{\delta}\left(\frac{\ell(c,d(1-\delta),1)}{\ell(a,b,1)}-1\right)\right]+O(1)\,, \tag{S152}$$

where ℓ is the function defined in Proposition S39.

So far, we have obtained effective ranks that have either $O(1)$ or $O(N)$ asymptotic behaviors as $N\to\infty$. We are going to prove $O(N^{1-\epsilon})$ asymptotic behaviors are also possible for all $\epsilon > 0$.

Corollary S42 (Hypergeometric decrease admits $O(N^{1-\epsilon})$ effective ranks)**.** *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 b_*, b^*, c_*, c^*, ζ_*, ζ^* such that $0\le b_*\le b^*$, $2\le c_*\le c^*$, $0<\zeta_*\le\zeta^*$ and such that for each N, the singular values $\sigma_1\ge\sigma_2\ge\cdots\sigma_N\ge 0$ of W_N satisfy*

$$\left(1-\frac{i-1}{N-1}\right)^{c^*-2}\left(1+\zeta^*\frac{i-1}{N-1}\right)^{-b^*} \le \frac{\sigma_i}{\sigma_1} \le \left(1-\frac{i-1}{N-1}\right)^{c_*-2}\left(1+\zeta_*\frac{i-1}{N-1}\right)^{-b_*} \qquad \forall\, i\in\{1,\ldots,N\}\,.$$

1. *If $c_*=c^*=N^\epsilon/d$ for some $d>0$ and $\epsilon\in(0,1]$, then as $N\to\infty$,*

$$\frac{d}{2}N^{1-\epsilon}+O(N^{1-2\epsilon}) \le \operatorname{srank}(W_N) \le 1+\frac{d}{2}N^{1-\epsilon}+O(N^{1-2\epsilon})\,,$$

$$d\,N^{1-\epsilon}+O(N^{1-2\epsilon}) \le \operatorname{nrank}(W_N) \le 1+d\,N^{1-\epsilon}+O(N^{1-2\epsilon})\,,$$

$$d\,N^{1-\epsilon}+O(N^{1-2\epsilon}) \le \operatorname{erank}(W_N) \le (1+d\,N^{1-\epsilon})e^{\frac{1}{1-\delta}}+O(\max(1,N^{1-2\epsilon}))\,,$$

where the last inequality holds for all $\delta\in(0,1)$.

2. *If $b_*>1$, $\zeta_*=\zeta^*=N^\epsilon/d$ for some $d>0$ and $\epsilon\in(0,1]$, and $b_*, b^*\notin\mathbb{Z}$, then*

$$\frac{d}{2b^*-1}N^{1-\epsilon}+O(N^{1-2\epsilon}) \le \operatorname{srank}(W_N) \le 1+\frac{d}{2b_*-1}N^{1-\epsilon}+O(N^{1-2\epsilon})\,,$$

$$\frac{d}{b^*-1}N^{1-\epsilon}+O(N^{1-2\epsilon}) \le \operatorname{nrank}(W_N) \le 1+\frac{d}{b_*-1}N^{1-\epsilon}+O(N^{1-2\epsilon})\,,$$

$$\frac{d}{b^*-1}N^{1-\epsilon}+O(N^{1-2\epsilon}) \le \operatorname{erank}(W_N) \le \left(1+\frac{d}{b_*-1}N^{1-\epsilon}\right)\exp\left(\frac{b^*-b_\delta}{\delta(b_\delta-1)}\right)+O(\max\{1,N^{1-b_\delta\epsilon},N^{1-2\epsilon}\})\,,$$

where the last inequality holds for all $\delta\in(0,1-1/b_*)$ and $b_\delta:=(1-\delta)b_*$.

3. If $b^* < 1$, $\zeta_* = \zeta^* = N^\epsilon/d$ for some $d > 0$, and $\epsilon \in (0,1]$, and $b_*, b^* \notin \mathbb{Z}$, then

$$g(2b^*, 2c^*-2, d)\, N^{1-2b^*\epsilon} + O\left(N^{1-(2b^*+1)\epsilon}\right) \ \le\ \mathrm{srank}(W_N)\ \le 1 + g(2b_*, 2c_*-2, d)\, N^{1-2b_*\epsilon} + O\left(N^{1-(2b_*+1)\epsilon}\right),$$
$$g(b^*, c^*, d)\, N^{1-b^*\epsilon} + O\left(N^{1-(b^*+1)\epsilon}\right) \ \le\ \mathrm{nrank}(W_N)\ \le\ 1 + g(b_*, c_*, d)\, N^{1-b_*\epsilon} + O\left(N^{1-(b_*+1)\epsilon}\right),$$
$$g(b^*, c^*, d)\, N^{1-b^*\epsilon} + O\left(N^{1-(b^*+1)\epsilon}\right) \ \le\ \mathrm{erank}(W_N)\ \le \left(1 + g(b_*, c_*, d)\, N^{1-b_*\epsilon}\right) e^{y(N)} + O\left(N^{1-\gamma\epsilon}\, e^{y(N)}\right),$$

where the last inequality holds for all $\delta \in (0, 1 + (1-2b^*)/b_*)$ with $b^* > 1/2$ and $b_* > 2b^* - 1$, $\gamma = 1 - 2(b^* - b_*) + \delta b_*$, and

$$g(b,c,d) = \frac{\Gamma(1-b)\Gamma(c-1)\, d^b}{\Gamma(c-b)}, \qquad y(N) = \frac{g((1-\delta)b_*, (1-\delta)c_* + 2\delta, d)}{\delta\, g(b^*, c^*, d)}\, N^{\epsilon(b^*-b_\delta)} - \frac{1}{\delta}. \tag{S153}$$

Proof. First, we use the following asymptotic expansion for $c \to \infty$ [140, (15.12.2)]:

$${}_2F_1(a,b;c;-\zeta) = 1 - \frac{ab}{c}\zeta + O(c^{-2}).$$

Hence, for $c = N^\epsilon/d$,

$$\frac{N-1}{c-1}\, {}_2F_1(a,b;c;-\zeta) = dN^{1-\epsilon}\left(1 + O(N^{-\epsilon})\right), \qquad \frac{N-1}{2c-3}\, {}_2F_1(a,b;c;-\zeta) = \frac{d}{2}N^{1-\epsilon}\left(1 + O(N^{-\epsilon})\right),$$

The substitution of the last equations into the bounds of Theorem S38 readily provides the desired inequalities for srank, nrank, and the lower bound of erank. For the upper bound of the erank, we need to get the asymptotics of

$$\mathcal{H} = \frac{1}{\delta}\left(\frac{c^*-1}{(1-\delta)c_* + 2\delta - 1}\, \frac{{}_2F_1(1, (1-\delta)b_*; (1-\delta)c_* + 2\delta; -\zeta_*)}{{}_2F_1(1, b^*; c^*; -\zeta^*)} - 1\right).$$

However,

$$\frac{{}_2F_1(1, (1-\delta)b_*; (1-\delta)c_* + 2\delta; -\zeta_*)}{{}_2F_1(1, b^*; c^*; -\zeta^*)} = 1 + O(N^{-\epsilon}), \qquad \frac{c^*-1}{(1-\delta)c_* + 2\delta - 1} = \frac{1}{1-\delta} + O(N^{-\epsilon}),$$

leading to the conclusion that $\mathcal{H} = \frac{1}{1-\delta} + O(N^{-\epsilon})$. Moreover,

$$\mathcal{P} = \frac{1}{N-1}\frac{(c^*-1)\rho(b^*, c^*, \zeta^*)}{\delta\ {}_2F_1(1, b_*; c_*; -\zeta_*)} = O(N^{\epsilon-1}),$$

from which we deduce that asymptotic expansions

$$\mathcal{H} + \mathcal{P} = \frac{1}{1-\delta} + O(N^{-\epsilon}) + O(N^{\epsilon-1}) \qquad \text{and} \qquad e^{\mathcal{H}+\mathcal{P}} = e^{\frac{1}{1-\delta}}\left(1 + O(N^{-\epsilon}) + O(N^{\epsilon-1})\right)$$

where the latter result holds for all $\delta \in (0,1)$. Hence,

$$\mathrm{erank}(W_N)\ \le \Big(1 + d\, N^{1-\epsilon} + O(N^{1-2\epsilon})\Big)e^{\frac{1}{1-\delta}}\Big(1 + O(N^{-\epsilon}) + O(N^{\epsilon-1})\Big) = e^{\frac{1}{1-\delta}} + d\, e^{\frac{1}{1-\delta}}\, N^{1-\epsilon} + O(\max(1, N^{1-2\epsilon}))$$

as expected.

Second, we use the following identity valid for $|z| > 1$ and $a - b \notin \mathbb{Z}$ [140, (15.8.2)]:

$${}_2F_1(a,b;c;z) = \frac{\Gamma(b-a)\Gamma(c)}{\Gamma(c-a)\Gamma(b)}\frac{1}{(-z)^a}\, {}_2F_1\left(a, a-c+1; a-b+1; \frac{1}{z}\right) + \frac{\Gamma(a-b)\Gamma(c)}{\Gamma(c-b)\Gamma(a)}\frac{1}{(-z)^b}\, {}_2F_1\left(b-c+1, b; b-a+1; \frac{1}{z}\right)$$

Thus, for $a = 1$ and $z = -\zeta$,

$${}_2F_1(1,b;c;-\zeta) = \frac{c-1}{b-1}\frac{1}{\zeta}\, {}_2F_1\left(1, 2-c; 2-b; -\frac{1}{\zeta}\right) + \frac{\Gamma(1-b)\Gamma(c)}{\Gamma(c-b)}\frac{1}{\zeta^b}\, {}_2F_1\left(b-c+1, b; b; -\frac{1}{\zeta}\right)$$

However, according to [140, (15.2.1)],

$${}_2F_1\left(\alpha, \beta; \gamma; -\frac{1}{\zeta}\right) = 1 - \frac{\alpha\beta}{1!\gamma}\frac{1}{\zeta} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)}\frac{1}{\zeta^2} - + \ldots = 1 + O(\zeta^{-1})$$

The last two results imply that as $\zeta \to \infty$ ($\zeta := \zeta^* = \zeta_* = N^\epsilon/d$),

$$ {}_2F_1(1, b; c; -\zeta) = f(b, c, \zeta) + O(\zeta^{-2}) + O(\zeta^{-b-1}), \qquad f(b, c, \zeta) = \frac{c-1}{b-1}\frac{1}{\zeta} + \frac{\Gamma(1-b)\Gamma(c)}{\Gamma(c-b)}\frac{1}{\zeta^b}. \tag{S154} $$

Substituting this result into Theorem S38, we get the following asymptotic expansion for $\zeta \to \infty$:

$$ \begin{aligned} \frac{N-1}{2c^*-3}\left(f(2b^*, 2c^*-2, \zeta) + O(\zeta^{-2}) + O(\zeta^{-2b^*-1})\right) &\leq \mathrm{srank}(W) \leq 1 + \frac{N-1}{2c_*-3}\left(f(2b_*, 2c_*-2, \zeta) + O(\zeta^{-2}) + O(\zeta^{-2b_*-1})\right), \\ \frac{N-1}{c^*-1}\left(f(b^*, c^*, \zeta) + O(\zeta^{-2}) + O(\zeta^{-b^*-1})\right) &\leq \mathrm{nrank}(W) \leq 1 + \frac{N-1}{c_*-1}\left(f(b_*, c_*, \zeta) + O(\zeta^{-2}) + O(\zeta^{-b_*-1})\right), \\ \frac{N-1}{c^*-1}\left(f(b^*, c^*, \zeta) + O(\zeta^{-2}) + O(\zeta^{-b^*-1})\right) &\leq \mathrm{erank}(W) \leq \left(1 + \frac{N-1}{c_*-1}\left(f(b_*, c_*, \zeta) + O(\zeta^{-2}) + O(\zeta^{-b_*-1})\right)\right) e^{\mathcal{H}+\mathcal{P}}. \end{aligned} $$

Setting $\zeta = N^\epsilon/d$ and simplifying the resulting expressions lead to the desired inequalities for srank, nrank, and the lower bound of erank. Regarding the upper bound of the erank, the use of Eq. (S154) allows us to write

$$ \frac{{}_2F_1(1, b_\delta; c_\delta; -\zeta_*)}{{}_2F_1(1, b^*; c^*; -\zeta^*)} = \frac{(c_\delta-1)(b^*-1)}{(b_\delta-1)(c^*-1)}\left[\frac{1 - r(b_\delta, c_\delta, d)\ N^{\epsilon(1-b_\delta)} + O(N^{-\epsilon}) + O(N^{-\epsilon b_\delta})}{1 - r(b^*, c^*, d)\ N^{\epsilon(1-b^*)} + O(N^{-\epsilon}) + O(N^{-\epsilon b^*})}\right], $$

where $b_\delta := (1-\delta)b_*$, $c_\delta := (1-\delta)c_* + 2\delta$, and

$$ r(b, c, d) = \frac{\Gamma(2-b)\,\Gamma(c-1)\,d^{b-1}}{\Gamma(c-b)}. $$

Moreover, setting $b_\delta > 1$ implies that

$$ \frac{{}_2F_1(1, b_\delta; c_\delta; -\zeta_*)}{{}_2F_1(1, b^*; c^*; -\zeta^*)} = \frac{(c_\delta-1)(b^*-1)}{(b_\delta-1)(c^*-1)}\left(1 + O\left(N^{\epsilon(1-b_\delta)}\right)\right)\left(1 + O\left(N^{\epsilon(1-b^*)}\right)\right) = \frac{(c_\delta-1)(b^*-1)}{(b_\delta-1)(c^*-1)} + O\left(N^{\epsilon(1-b_\delta)}\right) $$

along with

$$ \mathcal{H} = \frac{b^*-b_\delta}{\delta(b_\delta-1)} + O\left(N^{\epsilon(1-b_\delta)}\right) \qquad \text{and} \qquad \mathcal{P} = O\left(N^{\epsilon-1}\right). $$

Hence,

$$ e^{\mathcal{H}+\mathcal{P}} = \exp\left(\frac{b^*-b_\delta}{\delta(b_\delta-1)}\right)\left(1 + O\left(N^{\epsilon(1-b_\delta)}\right) + O\left(N^{\epsilon-1}\right)\right) $$

and the upper bound of the erank is

$$ \begin{aligned} \mathrm{erank}(W_N) &\leq \left(1 + \frac{d}{b_*-1}N^{1-\epsilon} + O(N^{1-2\epsilon})\right)\exp\left(\frac{b^*-b_\delta}{\delta(b_\delta-1)}\right)(1 + O(N^{\epsilon(1-b_\delta)}) + O(N^{\epsilon-1})) \\ &= \left(1 + \frac{d}{b_*-1}N^{1-\epsilon}\right)\exp\left(\frac{b^*-b_\delta}{\delta(b_\delta-1)}\right) + O(\max\{1, N^{1-\epsilon b_\delta}, N^{1-2\epsilon}\}). \end{aligned} $$

Third, setting $b^* < 1$ gives

$$ \frac{{}_2F_1(1, b_\delta; c_\delta; -\zeta_*)}{{}_2F_1(1, b^*; c^*; -\zeta^*)} = \frac{g(b_\delta, c_\delta, d)}{g(b^*, c^*, d)}\,N^{\epsilon(b^*-b_\delta)} + O(N^{\epsilon(2b^*-b_\delta-1)}), $$

where the function g is defined in Eq. (S153). This leads to

$$ \mathcal{H} = \frac{g(b_\delta, c_\delta, d)}{\delta\, g(b^*, c^*, d)}\,N^{\epsilon(b^*-b_\delta)} - \frac{1}{\delta} + O(N^{\epsilon(2b^*-b_\delta-1)}). $$

With $\mathcal{P} = O(N^{b^*\epsilon-1})$, we find

$$ e^{\mathcal{H}+\mathcal{P}} = e^{y(N)}\,e^{O(N^{\epsilon(2b^*-b_\delta-1)})} \qquad \text{with} \qquad y(N) = \frac{g(b_\delta, c_\delta, d)}{\delta\, g(b^*, c^*, d)}\,N^{\epsilon(b^*-b_\delta)} - \frac{1}{\delta} $$

If $2b^* - b_\delta - 1 < 0$, or equivalently, $b^* < \frac{b_\delta+1}{2}$, then

$$ \begin{aligned} \mathrm{erank}(W_N) &\leq \left(1 + g(b_*, c_*, d)\,N^{1-b_*\epsilon} + O\left(N^{1-(b_*+1)\epsilon}\right)\right)e^{y(N)}(1 + O(N^{\epsilon(2b^*-b_\delta-1)})) \\ &= \left(1 + g(b_*, c_*, d)\,N^{1-b_*\epsilon}\right)\,e^{y(N)} + O(N^{1-(1-2(b^*-b_*))\epsilon+\delta b_*\epsilon}\,e^{y(N)}) \end{aligned} $$

as desired. □

Remark S43. For the upper bound of the erank, we also note the following. In the case 2, additionally, if $b^* - b_* = \delta$

for some $\delta \in (0, 1 - 1/b_*)$, then

$$\exp\left(\frac{b^* - b_\delta}{\delta(b_\delta - 1)}\right) = \exp\left(\frac{b_* + 1}{b_* - 1}\right) + O(\delta)\,.$$

One can also simplify the upper bound for the erank in the case 3 by setting $b^* - b_* = \delta$ and $c^* - c_* = \gamma\delta$, by considering a small δ, and by using Stirling's formula for the gamma functions in $g(b_\delta, c_\delta, d)$.

It is worth emphasizing that the hypergeometric envelopes of the previous corollary, given their generality, not only admit $O(N^{1-\epsilon})$ growth of the effective ranks, but can also produce $O(1)$ and $O(N)$ growths. Indeed, when $b_* = b^* = 0$, one recovers special sub-linear to supra-linear decreasing envelopes included in Corollary S41, leading to $O(N)$ effective ranks. When $c_* = c^* = 2$ and $b_* = b^* \to \infty$, one gets exponentially decreasing envelopes as in Corollary S40, corresponding to $O(1)$ effective ranks. Finally, when $c_* = c^* = 2$ while b_* and b^* remain finite, one instead obtains power-law decreasing envelopes and it can be shown, using the asymptotics of the Hurwitz zeta function, that it leads to $O(N^{1-\epsilon})$ effective ranks.

We have thus proved that different choices for ψ_* and ψ^* can induce very distinct asymptotic behaviors of the above-mentioned effective ranks as $N \to \infty$. Figure S4 depicts these findings by showing different singular-value envelopes, leading to three different classes of maximum growth of nrank as N becomes large: linear $O(N)$, sub-linear $O(N^{1-\epsilon})$ with $0 < \epsilon < 1$, and constant $O(1)$.

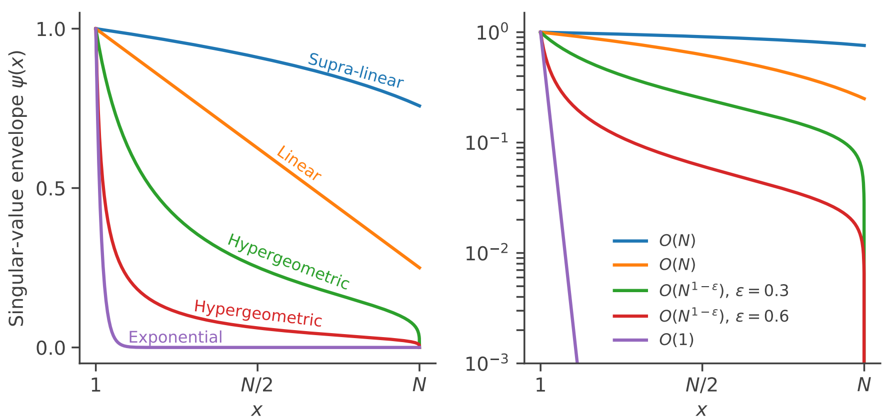

Fig. S4: Typical singular-value envelopes $\psi(x)$, describing the decreasing behavior of the normalized singular values σ_i/σ_1, vs. the continuous variable x, interpolating between the indices i, and their impact on the asymptotics of srank, nrank, and erank (bottom-right panel). The scale of values for $\psi(x)$ is linear on the left while it is logarithmic on the right. From top to bottom, the decreasing functions are: $\left(1 - a\frac{x-1}{N-1}\right)^b$ with $a = \frac{3}{4}$, $b = \frac{1}{5}$ (blue), and $b = 1$ (orange); $\left(1 - \frac{x-1}{N-1}\right)^{c-2} / \left(1 + \zeta\frac{x-1}{N-1}\right)^b$ with $b = \frac{3}{4}$, $c = 2\frac{1}{4}$, $\zeta = N^{0.3}$ (green), and $\zeta = N^{0.6}$ (red); ω^{x-1} with $\omega = 0.97$ (purple). For all cases, $N = 2000$.

As previously mentioned, the hypergeometric case generalizes several types of decrease, including the power-law decrease ($c_* = c^* = 2$). The latter has been observed in the adjacency spectrum of scale-free random graphs [69] and in the eigenspectra of covariance matrices in various settings, including fractional Brownian motion [141], EEG time series [142], neuronal activity in the mouse [143, 144] and macaque [145] visual cortex.

In their seminal work, Stringer et al. [143] examined the eigenspectrum of the covariance matrix derived from signals of large populations of neurons in the visual cortex of awake mice viewing natural images. They argued that the evoked neuronal population activity in this context is "high-dimensional". Their conception of high dimensionality is anchored in sophisticated theoretical findings, including methods from functional data analysis and fractal geometry. As they noted [143, Supplementary information p.7]: *These results demonstrate that unless eigenspectra decay faster than* n^{-1} *, population codes are pathological, either exhibiting discontinuous responses, or infinite population variance. Furthermore, for stimuli drawn from a set of manifold dimension d, codes with eigenspectra decaying slower than* $n^{-1-2/d}$ *are also pathological, displaying infinite variance of the code's derivative and fractal geometry of the response manifold. We conclude that our experimental observations of eigenspectrum decay only just faster than* $n^{-1-2/d}$ *indicate a neural code that is as high-dimensional as possible before hitting the regime where these pathological conditions must occur.* Therefore, the dimensionality of the neural code is deemed "high" when the eigenvalues of the covariance matrix decrease as slow as possible, nearing the threshold indicative of pathological responses, and "low" when the decrease is faster, significantly distanced from this critical threshold.

However, no measure of dimension is used to quantify the decrease of the eigenvalues. In fact, as shown below, the decrease of the covariance-matrix eigenvalues implied by the above power-law is fast enough to lead to effective ranks of order strictly less than $O(N)$, where N corresponds to the number of neurons, thus suggesting low dimensionality from our perspective. Indeed, let us assume that the visual stimuli's manifold dimension $d_{\text{stim}} > 2$ and that the n-th eigenvalue of the covariance matrix is less than $n^{-1-2/d_{\text{stim}}}$ as described in [143], implying that its n-th singular value is bounded above by $n^{-1/2-1/d_{\text{stim}}}$. Moreover, let us assume without loss of generality that $\sigma_1 = 1$. This scenario aligns with the third case of Corollary S42 for $b_* = 1/2 + 1/d_{\text{stim}} < 1$, $c_* = 2$, and $\zeta_* = N - 1 \approx N$ (i.e., $d = 1$, $\epsilon = 1$). We can deduce from this that srank and nrank asymptotically grow as $1 + O(N^{-2/d_{\text{stim}}})$ and $O(N^{1/2-1/d_{\text{stim}}})$, respectively. This presents an intriguing case where at least two effective ranks exhibit completely different asymptotic behaviors. Yet, in this power-law scenario, both the stable rank to dimension ratio and nuclear rank to dimension ratio tend to zero as N grows to infinity.

3. Impact of matrix density on the stable rank

Let us now derive some intuitive inequalities for the stable rank of graphs based on inequalities for the weight matrix.

Lemma S44. *Let W be a $N \times N$ matrix. Then the Frobenius norm of W is upper bounded as*

$$\|W\|_F \leq N \max_{i,j} |W_{ij}| \,. \tag{S155}$$

Moreover, the spectral norm of W is lower bounded as

$$\|W\|_2 \geq \max\left\{ \max_i \|\boldsymbol{r}_i\|,\, \max_j \|\boldsymbol{c}_j\|,\, \frac{1}{\sqrt{N}}\|\boldsymbol{k}^{\text{in}}\|,\, \frac{1}{\sqrt{N}}\|\boldsymbol{k}^{\text{out}}\| \right\}, \tag{S156}$$

where $\boldsymbol{r}_i$ and $\boldsymbol{c}_j$ respectively denote the i-th row and j-th column of W while $\boldsymbol{k}^{\text{in}} = W\boldsymbol{1}$ and $\boldsymbol{k}^{\text{out}} = W^\top \boldsymbol{1}$.

Proof. The first inequality trivially follows from the definition of the Frobenius norm:

$$\|W\|_F = \sqrt{\sum_{i,j} W_{ij}^2} \leq \sqrt{\sum_{i,j} W_{\max}^2} = \sqrt{N^2 W_{\max}^2} = N W_{\max} \,, \qquad W_{\max} = \max_{i,j} |W_{ij}| \,.$$

The second inequality is the maximum between four lower bounds. To derive them, we start with the definition

$$\|W\|_2 = \max_{\|\boldsymbol{x}\|=1} \|W\boldsymbol{x}\| \,,$$

which implies that $\|W\|_2 \geq \|W\boldsymbol{x}\|$ for any $\boldsymbol{x}$ such that $\|\boldsymbol{x}\| = 1$. Choosing $\boldsymbol{x} = \boldsymbol{e}_j$, the j-th unit vector, leads to the inequality $\|W\|_2 \geq \|\boldsymbol{c}_j\|$. But this inequality holds all j, so we can combine all the inequalities and infer that

$$\|W\|_2 \geq \max_j \|\boldsymbol{c}_j\|.$$

Now, because the spectral norm is invariant under matrix transposition, we also know that $\|W\|_2 \geq \left\|W^\top \boldsymbol{x}\right\|$ for any $\boldsymbol{x}$ such that $\|\boldsymbol{x}\| = 1$. Setting once again $\boldsymbol{x} = \boldsymbol{e}_i$ for all i, we conclude that

$$\|W\|_2 \geq \max_i \|\boldsymbol{r}_i\|.$$

Choosing $\boldsymbol{x} = \boldsymbol{1}/\sqrt{N}$ in $\|W\|_2 \geq \|W\boldsymbol{x}\|$ and in $\|W\|_2 \geq \left\|W^\top \boldsymbol{x}\right\|$ readily provides the two other lower bounds. □

Proposition S45. *Let W be the adjacency matrix of a directed graph of N vertices and M edges. Moreover, let $k_{\max}$ be the maximum among all ingoing and outgoing degrees of the graph. Then,*

$$\text{srank}(W) \leq \frac{M}{k_{\max}} \,. \tag{S157}$$

Proof. We first note that when W is an adjacency matrix, all its elements are either 0 or 1, which implies that its Frobenius norm squared is exactly equal to M. Indeed,

$$\|W\|_F^2 = \sum_{i,j} W_{ij}^2 = \sum_{\substack{(j,i) \\ j \to i}} 1 = M. \tag{S158}$$

Moreover, we know from the previous lemma that the spectral norm squared is bounded by the degrees:

$$\|W\|_2^2 \geq \max\left\{ \max_i \sum_j W_{ij}^2,\, \max_j \sum_i W_{ij}^2 \right\} = \max\left\{ \max_i \sum_j W_{ij},\, \max_j \sum_i W_{ij} \right\} = \max\left\{ \max_i k_i^{\text{in}},\, \max_j k_j^{\text{out}} \right\} = k_{\max} \,.$$

Thus, $\text{srank}(W) = \|W\|_F^2 / \|W\|_2^2 \leq M/k_{\max}$, as expected. □

In dense directed graphs of N vertices, the number of edges M typically scales as N^2 while the maximum degree scales as N. The previous proposition thus implies that the stable rank is of order $O(N)$ for such graphs. A slightly different scaling law exists for sparse graphs. Indeed, if $M = O(N^{2-\varepsilon})$ and $k_{\max} = O(N^{1-\epsilon})$ for some $\varepsilon > \epsilon > 0$, then the stable rank is of order $O(N^{1+\epsilon-\varepsilon}) = o(N)$. As shown in next proposition, similar scaling behaviors emerge when considering general square matrices, which are relevant for studying signed weighted directed graphs.

Proposition S46. *Let $p \geq 0$ and $0 < \alpha \leq \beta$. Let W be a $N \times N$ matrix such that*

$$\alpha N^{-p} \leq W_{ij}^2 \leq \beta N^{-p} \tag{S159}$$

for all $1 \leq i, j \leq N$. Then, the stable rank satisfies the inequality

$$\mathrm{srank}(W) \leq \alpha^{-1}\beta\, N\,. \tag{S160}$$

More generally, if the maximum number of nonzero elements in a row or in a column of W is γN, the total number of nonzero elements of W is δN^2, and all these nonzero elements satisfy inequality (S159), then

$$\mathrm{srank}(W) \leq \alpha^{-1}\beta\gamma^{-1}\delta\, N\,. \tag{S161}$$

Proof. We use Lemma S44 and proceed essentially as for Proposition S45. □

Sparse $N \times N$ matrices are characterized by a total number of nonzero elements of order strictly less than N^2 and a maximum number of nonzero elements in each row or column of order strictly less than N. In the last proposition, this situation corresponds to the case where $\gamma = \bar{\gamma}\, N^{-\epsilon}$ and $\delta = \bar{\delta} N^{-\varepsilon}$ for some $\varepsilon > \epsilon > 0$, which implies that once again, $\mathrm{srank} = O(N^{1+\epsilon-\varepsilon}) = o(N)$. A typical sparse matrix has $\varepsilon = 2\epsilon$, leading to a stable rank scaling as $O(N^{1-\varepsilon})$, which tends to $O(1)$ when considering the ultra-sparse case $\epsilon = 1$. In words, the stable rank of (signed weighted directed) graphs having N vertices increases at most linearly with N and sparsity makes the increase become sub-linear. This means that sparse graphs are characterized by a ratio srank/N that goes to zero as N grows, obviously corresponding to a low effective rank.

D. Directed network centrality measures

For a directed network, the eigenvalues and the eigenvectors of its matrix representation will generally be complex and one have to adapt the usual approach to define a centrality. A natural way of doing that is to use the SVD of the directed network, which provides two vertex centrality measures: the authority centrality (dominant left singular vector u_1) and the hub centrality (dominant right singular vector v_1) [146, 147], as illustrated in Fig. S5. This remark guided us in choosing the observables of the reduced dynamics and it can be used to give an interpretation to the different terms and equations involved when applying Theorem S52 with the reduction matrix being the right singular vectors. Note, however, that for signed networks (described by matrices with negative values), these centrality measures may lead to ambiguities, since the first left and right singular vectors generally have negative values (Perron-Frobenius theorem [71, Theorem 38] doesn't apply).

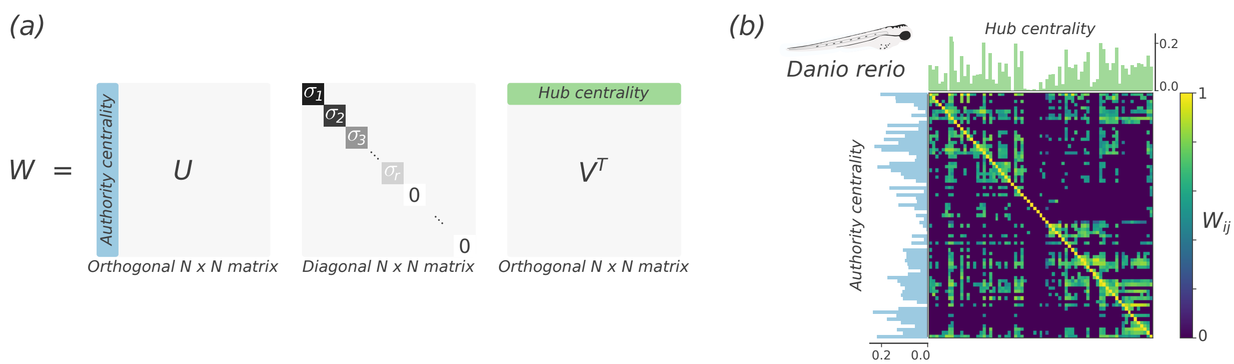

Fig. S5: (a) Authority and hub centralities are provided by the elements of the dominant left and right singular vectors, respectively. (b) Centralities for the mesoscopic connectome of the larval zebrafish with $N = 71$ communities with added self-loops (modified from Ref. [148]).

E. Adaptive networks

Complex systems are not only characterized by their nonlinear dynamics and network structure, but also by their capacity to adapt themselves to environmental changes [149]. The effective rank of a complex network should thus be expected to change according to time. We performed a preliminary investigation of this phenomenon by extracting the effective rank of the *C. elegans* connectome at different stages of its maturation [150] as shown in Fig S6. We observed that the stable rank decreases with age. More work should be done on this subject to verify if this decrease is significant and to determine the biological meaning of an effective rank decrease with maturation.

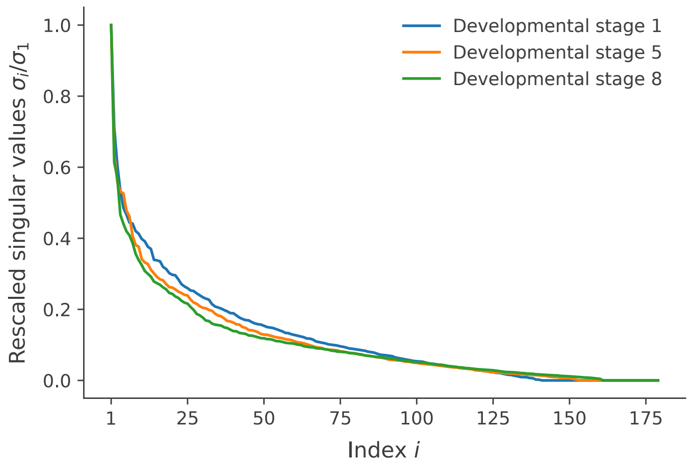

Fig. S6: Singular values of the matrices describing the connectivity of the *C. elegans* brain at different maturation stages. The stable ranks are 21.6 (developmental stage 1), 19.7 (developmental stage 5), 18.5 (developmental stage 8).

In Ref. [128], the authors numerically show that training a neural network decreases the stable rank, which is somewhat in line with what we observe in the latter biological example.

F. SVD for dynamical systems

The applications of SVD for dynamical systems is very broad, especially in engineering and linear control systems [40]. SVD is also generalized for nonlinear operators [151] and it is even possible to perform a quasi-optimal low-rank approximation for matrix dynamics with time-evolving matrices $U, \Sigma, V^\top$ [152], which could have interesting applications in the study of temporal networks [153]. As illustrated in the paper, one can also leverage the power of SVD in the dimension reduction of dynamical systems on networks. As explained in Ref. [154, Appendix C], it can be very hard to choose adequately the reduction matrix M. Having real nonnegative singular values and real singular vectors is an advantage when it comes to interpreting the spectra and to define interpretable observables for the dynamics (as opposed to eigenvalue decomposition for general real matrices, which can raise the problem of dealing with complex reduction matrices and create complex reduced dynamics for an initially real dynamics [155, p.145][154]). In the following section, we give details about the dimension reduction of complex systems and especially, in subsection III D, we show how to use the salient properties of SVD to get insights on the low-rank hypothesis of complex systems.

III. DIMENSION REDUCTION OF COMPLEX SYSTEMS

Dimension reduction of high-dimensional dynamics is a powerful technique to get analytical and numerical insights on complex systems. For instance, it helps predict the onset of explosive phenomena [156] or getting suitable observable to assess the controllability of the system [157]. The range of applications of dimension-reduction techniques is therefore very broad—ranging from statistical physics and chemistry to finance and neuroscience— and the methods substantially differ along with the terminology: dimension reduction [12, 158], coarse graining [159–161], reduced-order model [4], model reduction [162], lumping [163, 164] [165, Section 2.4], compression [166], pruning [167], dominance analysis [168], variable or state aggregation [169], etc. Many useful dimension-reduction techniques remain unused for complex systems which may be a consequence of this great diversity of terminologies. In this section, we give details about dimension reduction of ordinary differential equations from its more general aspects to the specific ones used in the paper.

A. Notation and generalities on dimension reduction

Consider the following notation for the complete dynamical systems:

- $x \in \mathbb{R}^N$ is a state of the system;
- $t \in [0, \infty)$ denotes time;
- $\phi : [0, \infty) \times \mathbb{R}^N \mapsto \mathbb{R}^N$ is the flow;
- $x : [0, \infty) \to \mathbb{R}^N$ is the trajectory (note the abuse of notation with the state);
- $f : \mathbb{R}^N \to \mathbb{R}^N$ is the vector field, assumed to be continuously differentiable;
- $x_0 = x(0)$ is the initial condition;
- $\dot{x} = f(x)$ is the complete dynamics, or more explicitly,

$$\begin{pmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_N \end{pmatrix} = \begin{pmatrix} f_1(x_1, ..., x_N) \\ \vdots \\ f_N(x_1, ..., x_N). \end{pmatrix}$$

Consider the following notation for the reduced dynamical system:

- $R : \mathbb{R}^N \mapsto \mathbb{R}^n$ with $n < N$ is called the reduction function or a vectorial observable;
- $R = (R_1, ..., R_n)$ where $R_\mu : \mathbb{R}^N \mapsto \mathbb{R}$ is the μ-th observable;
- $X = R(x) \in \mathbb{R}^n$ is a reduced state;
- $\Phi : [0, \infty) \times \mathbb{R}^n \mapsto \mathbb{R}^n$ is the reduced flow;
- $X = R \circ x : [0, \infty) \to \mathbb{R}^n$ is the reduced trajectory (note the abuse of notation with the reduced state);
- $F : \mathbb{R}^n \to \mathbb{R}^n$ is the reduced vector field, assumed to be continuously differentiable;
- $X_0 = R(x_0) = (R \circ x)(0)$ is the initial condition;
- $\dot{X} = F(X)$ is the reduced dynamics.

The logic behind the notation is that the "microscopic" objects are in lowercase and the "macroscopic" objects are in uppercase, except for N and n which denote some high dimension and a lower dimension respectively. Latin indices are used for these microscopic objects, while Greek indices are used for the macroscopic objects. With this notation, we now define what we mean by exact dimension reduction, in a similar spirit as Ref. [164], but avoiding the subtleties in the characteristics of the reduction function R.

Definition S47. The function $R : \mathbb{R}^N \mapsto \mathbb{R}^n$ induces an *exact dimension reduction* of the dynamics

$$\dot{x} = f(x) \tag{S162}$$

if there exists a vector field $F : \mathbb{R}^n \to \mathbb{R}^n$ such that for all solutions $x(t)$ of Eq. (S162), the reduced trajectory

$$X = R \circ x : [0, \infty) \to \mathbb{R}^n \tag{S163}$$

obeys the differential equation

$$\dot{X} = F(X)\,. \tag{S164}$$

The pair of functions (R, F) thus characterizes a dimension reduction, where the goal is to *close* the differential equation for X in terms of X solely. Dimension reduction can also be seen as a special commutation relation of the vector fields and the flows.

Theorem S48. *The following statements are equivalent:*

1. *the dimension reduction is exact;*
2. *the general compatibility equation*

$$\mathcal{U}[R] = J_R \circ f = F \circ R, \tag{S165}$$

holds where $\mathcal{U}$ is the Koopman operator generator and J_R is the Jacobian matrix of R;

3. *the complete flow ϕ_t and the reduced flows Φ_t commutes with R such that*

$$R \circ \phi_t = \Phi_t \circ R. \tag{S166}$$

Proof.
(1. $\Leftarrow$ 2.) By definition, $X = R \circ x$ and by assumption, $J_R f = F \circ R$. Then, the time derivative of X (the generator of the Koopman operator) is

$$\dot{X} = \frac{\mathrm{d}(R \circ x)}{\mathrm{d}t} = \mathcal{U}[R] \circ x = J_R \circ f \circ x = F \circ R \circ x = F \circ X, \tag{S167}$$

which is the definition of an exact dimension reduction.
(1. $\Rightarrow$ 2.) Similarly, using the time derivative of $X = R \circ x$ again, we have

$$\dot{X} = \mathcal{U}[R] \circ x = J_R \circ f \circ x. \tag{S168}$$

But the dimension reduction is exact and $\dot{X} = F \circ X = F \circ R \circ x$ holds. Then, by comparison, it is sufficient to have $\mathcal{U}[R] \circ x = J_R \circ f = F \circ R$.
(1. $\Leftrightarrow$ 3.) On the one hand, the solution of $\dot{x} = f(x)$ is $x(t) = \phi_t(x(0))$ and thus, the exact evolution of $X(t)$ is given by $X(t) = R \circ \phi_t \circ x(0)$. On the other hand, the solution to $\dot{X} = F(X)$ with $X(0) = R(x(0))$ is $X(t) = \Phi_t(X(0)) = \Phi_t \circ R \circ x(0)$. The comparison gives the desired result. □

Since we have commutation relations, there is a clear picture with commutative diagrams. In particular, statement 3. tells us that that we have an exact dimension reduction if there is a commutative diagram such that

$$\begin{tikzcd} \mathbb{R}^N \arrow[r, "\phi_t"] \arrow[d, "R"'] \arrow[dr, dotted, shift left, "R \circ \phi_t"] \arrow[dr, dotted, shift right, "\Phi_t \circ R"'] & \mathbb{R}^N \arrow[d, "R"] \\ \mathbb{R}^n \arrow[r, "\Phi_t"'] & \mathbb{R}^n \end{tikzcd} \tag{S169}$$

In the article and the rest of the Supplementary information, we focus on the case where R is a *linear transformation*, which greatly simplifies the analysis and gives access to a whole range of notions and tools from linear algebra. Let us thus assume that $X = R(x) = Mx$ where M is a $n \times N$ matrix, called the reduction matrix [154] (or lumping matrix [163, 170]). Then, $J_R = M$ and condition (S165) for closure states that for an exact dimension reduction, the complete and reduced vector fields must commute with M:

$$\begin{tikzcd} \mathbb{R}^N \arrow[r, "f"] \arrow[d, "M"'] \arrow[dr, dotted, shift left, "M \circ f"] \arrow[dr, dotted, shift right, "F \circ M"'] & \mathbb{R}^N \arrow[d, "M"] \\ \mathbb{R}^n \arrow[r, "F"'] & \mathbb{R}^n \end{tikzcd} \tag{S170}$$

where we have made a slight abuse of notation, using the same symbol for the matrix and the linear transformation $M : x \mapsto Mx$, that we will use again in the document. Note that the latter scheme is related to the notions of C^k-equivalent and C^k-conjugate vector fields defined in Ref. [171, p.190 and p.191]. In subsection III D [Definition S51], we introduce the alignment error which is directly defined from the compatibility equation $M \circ f = F \circ M$ and we will find a bound on it.

Remark S49. In our work, we consider that the network of the system is already known (or could be known experimentally) and the dynamics is described by a given theoretical model, but the time series/functional data (trajectories) are unknown. This is the ideal setting for determining how the low effective rank of the weight matrix W can affect the evolution of the state of the whole system, starting with arbitrary initial conditions, since no limitation in our analysis can be induced by the finite number of observed time series or their finite length.

Given a reduction matrix M, a projector can always be defined as

$$P = M^+ M\,, \tag{S171}$$

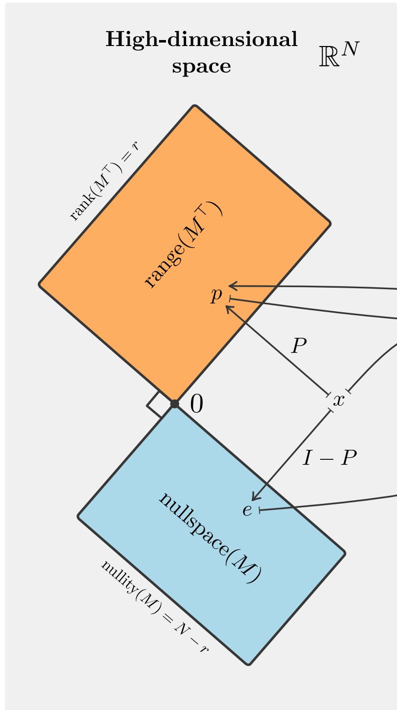

Fig. S7: Schematization of dimension reduction associated with the reduction matrix M, the corresponding projector P, and the induced vector subspaces.

where M^+ is the Moore-Penrose pseudo-inverse of M. Under this linear setup, the dimension reduction can be seen as a projection of the elements x of the high-dimensional space unto a low-dimensional space with elements X. This situation as well as the four natural vector subspaces induced by M are illustrated in Fig. S7.

In general, it is far from simple to solve the compatibility equations $M \circ f = F \circ M$ for F and M. Even when f and F are linear transformations, respectively encoded by the $N \times N$ matrix W and the $n \times n$ matrix $\mathcal{W}$, the condition $M \circ f = F \circ M$ takes the form of the compatibility equation [154]

$$MW = \mathcal{W}M \tag{S172}$$

which is in fact a system of coupled quadratic equations in the elements of $\mathcal{W}$ and M that cannot always be solved analytically. However, for a fixed M, one can find a unique optimal reduced matrix $\mathcal{W}$.

Theorem S50 (Ref. [154])**.** *Let M and W be respectively of size $n \times N$ and $N \times N$ with $n < N$. Then, the compatibility equation $\mathcal{W}M = MW$ has a solution for $\mathcal{W}$ if and only if*

$$MWM^+M = MW \tag{S173}$$

where M^+ is the Moore-Penrose pseudoinverse of M, in which case the solution is

$$\mathcal{W} = MWM^+ + Y - YMM^+, \tag{S174}$$

where Y is an arbitrary $n \times n$ matrix. If $\operatorname{rank} M = n$, then there is at most one solution to the compatibility equation, i.e.,

$$\mathcal{W} = MWM^+. \tag{S175}$$

For any M, this solution minimizes $\|\mathcal{W}M - MW\|_F$ with error $\|MW(I - M^+M)\|_F$.

Proof. From Penrose 1955 [44, Theorem 2], a necessary and sufficient condition for the equation $AXB = C$ to have a solution is $AA^+CB^+B = C$, in which case the general solution is $X = A^+CB^+ + Y - A^+AYBB^+$, where Y is arbitrary. Set $A = I$, $X = \mathcal{W}$, $B = M$, $C = MW$ in the latter equations and the result in Eq. (S174) follows. If $\operatorname{rank} M = n$, then the n rows of M are linearly independent. This implies that $MM^+ = I$ and

$$\mathcal{W} = MWM^+ + Y - Y = MWM^+,$$

which does not depend on the arbitrary matrix Y anymore. It is thus the only possible solution.

Finally, it is well known, at least since the least-squares theorem of Penrose in 1956 [172], that

$$\underset{U\in\mathbb{R}^{k\times\ell}}{\arg\min}\,\|UA-V\|_F = VA^+ \quad \text{and} \quad \min_{U\in\mathbb{R}^{k\times\ell}}\|UA-V\|_F = \|V(I-A^+A)\|_F,$$

for matrices $V\in\mathbb{R}^{k\times m}$ and $A\in\mathbb{R}^{\ell\times m}$. Setting $A=M$, $U=\mathcal{W}$, and $V=MW$ implies that $\mathcal{W}=MWM^+$ minimizes $\|\mathcal{W}M-MW\|_F$ with error $\|MW(I-M^+M)\|_F$. □

As it will be discussed in another paper, the first preliminary results on these compatibility equations seems to go back to 1969 in chemistry [163] and for a fixed $\mathcal{W}$, the compatibility equations are homogeneous Sylvester equations (1884) [173].

In the next section, we provide a way to find an optimal reduced vector field F given a reduction matrix M, thus generalizing the idea behind Theorem S50.

B. Least-square optimal vector field

Low-dimensional dynamical systems can be obtained from an optimization problem, where some error is minimized under a set of constraints [174] in order to preserve the salient properties of the original high-dimensional system. For dynamical systems, a natural optimization variable is the reduced vector field F itself, which is chosen to represent approximately the complete vector field f. Yet, it is rather puzzling to find how the different vector field errors are related to each other and which one can be minimized analytically. We found that there was a useful scheme that helps solve this puzzle. Recalling the definitions of subsection III A, we introduce the following diagram for dimension reduction of dynamical systems:

$$\begin{tikzcd}
x \arrow[rrr, mapsto, "f"] \arrow[drr, mapsto, "P"] \arrow[dddr, mapsto, bend right, "M"'] & & & f(x) \arrow[ddr, mapsto, "M"] \arrow[drr, dash, dashed, "\varepsilon(x)"] & & \\
& & Px \arrow[rr, mapsto, "f"] \arrow[ddl, mapsto, bend right, "M"'] & & f(Px) \arrow[r, dash, dashed, "\varepsilon'(x)"] \arrow[ddr, mapsto, "M"] & M^+F(X) \\
& & & & Mf(x) \arrow[d, dash, dashed, "\mathcal{E}(x)"] & \\
& X \arrow[rrr, mapsto, "F"'] \arrow[uur, mapsto, bend right, "M^+"'] & & & F(X) \arrow[uur, mapsto, "M^+"] \arrow[r, dash, dashed, "\mathcal{E}'(x)"'] & Mf(Px)
\end{tikzcd}, \tag{S176}$$

where $P=M^+M$ and the dashed lines represent root-mean-square errors (RMSE) between adjacent vector fields, i.e., different alignment errors as defined below [see also Fig. 3 for an illustration of $\mathcal{E}(x)$].

Definition S51. Let f be a complete vector field in $\mathbb{R}^N$, F be a reduced vector field in $\mathbb{R}^n$, and M be the $n\times N$ reduction matrix. At $x\in\mathbb{R}^N$, the *alignment error* ...

- ... in $\mathbb{R}^N$ is the RMSE between the vector fields f and $M^+\circ F\circ M$, i.e.,

$$\varepsilon(x) = \frac{1}{\sqrt{N}}\|f(x)-M^+F(Mx)\|\,; \tag{S177}$$

- ... in $\mathbb{R}^n$ is the RMSE between the vector field $M\circ f$ and $F\circ M$, i.e.,

$$\mathcal{E}(x) = \frac{1}{\sqrt{n}}\|Mf(x)-F(Mx)\|\,, \tag{S178}$$

where $\|\ \|$ is the Euclidean vector norm.

By applying the definition of alignment errors on the projected complete vector field $f\circ P$ instead of f only, we have defined

$$\varepsilon'(x) = \frac{1}{\sqrt{N}}\|f(Px)-M^+F(Mx)\| \tag{S179}$$

and

$$\mathcal{E}'(x) = \frac{1}{\sqrt{n}}\|Mf(Px) - F(Mx)\| \tag{S180}$$

in Diagram S176. In principle, the alignment error $\mathcal{E}(x)$ in $\mathbb{R}^n$ is to be minimized in order to be as close as possible to an exact dimension reduction [Definition S47, Theorem S48, and Diagram S170], but this is far from a simple task. However, the alignment error $\varepsilon'(x)$ can be directly minimized using least squares which has for consequence that the alignment error $\mathcal{E}'(x)$ in $\mathbb{R}^n$ is exactly 0, as shown in the following theorem.

Theorem S52. *Let f be a complete vector field in $\mathbb{R}^N$, F be a reduced vector field in $\mathbb{R}^n$, and M be a $n\times N$ reduction matrix. The vector field of the reduced dynamics*

$$\dot{X} = Mf(M^+X) \tag{S181}$$

is optimal in the sense that it minimizes the alignment error $\varepsilon'(x)$ in $\mathbb{R}^N$, i.e.,

$$F^*(X) = \underset{\substack{F(X)\in\mathbb{R}^n\\ X=Mx}}{\arg\min}\, \|f(Px) - M^+F(X)\| = Mf(M^+X). \tag{S182}$$

Consequently, the alignment error $\mathcal{E}'(x)$ in $\mathbb{R}^n$ is 0.

Proof. Let $v\in\mathbb{R}^k$ and $A\in\mathbb{R}^{k\times\ell}$. Then, using least squares (particular case of Penrose [172]) implies that

$$\underset{u\in\mathbb{R}^\ell}{\arg\min}\, \|v - Au\| = A^+v. \tag{S183}$$

Setting $A = M^+$, $u = F(X)$, $v = f(Px) = f(M^+Mx) = f(M^+X)$ readily yields the result. Since $F^*(X) = Mf(M^+X)$ and $Px = M^+X$, we obviously have $\|Mf(Px) - F^*(X)\| = 0$. ☐

Remark S53.

- Minimizing $\varepsilon'(x)$ does not tell much about the alignment error $\mathcal{E}(x)$ of interest. Yet, in subsection III D, we find that using the ensuing vector field from the minimization of $\varepsilon'(x)$ allows obtaining an upper bound on $\mathcal{E}(x)$.
- Recalling the optimal solution MWM^+ for the compatibility equation $MW = \mathcal{W}M$ in Theorem S50, we observe that we now have an optimal solution (involving a nonlinear vector field) $M\circ f\circ M^+$ for the compatibility equation $M\circ f = F\circ M$ that boils down to the previous linear solution when $f = W$ and $F = \mathcal{W}$.
- When we set $n = N$, we could expect the "reduced" vector field F to be equivalent in some way to the complete vector field. In fact, if $f:\mathbb{R}^N\to\mathbb{R}^N$ is a $\mathcal{C}^1(\mathbb{R}^N)$ vector field, the vector fields f and $F = M\circ f\circ M^+ = M\circ f\circ M^{-1}$ are $\mathcal{C}^1$-conjugate on $\mathbb{R}^N$ [171, p.191], which is straightforward to observe from the form of F itself where M is the $\mathcal{C}^1$-diffeomorphism.
- To the authors' knowledge, even if the vector field in Eq. (S182) is known at least since 1989 [175], the result hasn't been stated and proved clearly, simply, and in a general way for dynamical systems described by a set of differential equations. One can find many papers on the method (e.g., in fluid mechanics and chemistry) [175–178] and especially, on a similar-looking technique for time series which is also loosely [179, 180] called Galerkin projection or Petrov-Galerkin method [4, 40, 177, 181]. In our paper, we recall that it is implicitly assumed that we do not have access to the time series, only the initial vector field with the network is known.
- In principle, there is a whole world of objective functions that could be used for the optimization problem. Other constraints and regularization terms could also be added to satisfy the modeler's restrictions. This is a promising avenue to be further explored in the future to obtain optimal reduced dynamical systems.

Let us now apply the latter theorem to one of the most influential models in neuroscience, the Wilson-Cowan model [182–185] [186, Chap. 11].

Example S54 (Neuroscience)**.** Consider a system of $N = N_E + N_I$ neurons (or neuronal population) with N_E excitatory neurons and N_I inhibitory neurons. Let E_e (resp. I_i) be the time-averaged firing rate of the e-th excitatory neuron for $e\in\mathcal{E} = \{1,...,N_E\}$ (resp. i-th inhibitory population for $i\in\mathcal{I} = \{N_E+1,...,N\}$). The Wilson-Cowan model [182] describes the time evolution of the firing rates as

$$\dot{E}_e = -d_eE_e + (1 - aE_e)\;\mathcal{S}[b(\textstyle\sum_{e'=1}^{N_E}W_{ee'}E_{e'} + \sum_{i'=N_E+1}^{N}W_{ei'}I_{i'} - c_e)] \tag{S184}$$

$$\dot{I}_i = -d_iI_i + (1 - aI_i)\;\mathcal{S}[b(\textstyle\sum_{e'=1}^{N_E}W_{ie'}E_{e'} + \sum_{i'=N_E+1}^{N}W_{ii'}I_{i'} - c_i)], \tag{S185}$$

where d_y is the inverse time constant and a is related to the refractory period. Moreover, for all $i,i'\in\mathcal{I}$ and $e,e'\in\mathcal{E}$, $W_{ee'}\geq 0$, $W_{ie'}\geq 0$, $W_{ei'}\leq 0$, $W_{ii'}\leq 0$, and

$$\mathcal{S}[b(z - c)] = \frac{1}{1 + e^{-b(z-c)}} \tag{S186}$$

is the logistic function with b being its steepness and c being its midpoint or physically, an external input. By defining

$$(x_1, ..., x_N) := (E_1, ..., E_{N_E}, I_{N_E+1}, ..., I_N)^\top, \tag{S187}$$

we get a concise form of the model [185, Eq. (11)]:

$$\dot{x}_j = -d_j x_j + (1 - a x_j)\mathcal{S}[b(\gamma\, y_j - c)], \quad \forall\, j \in \{1, ..., N\}, \tag{S188}$$

where $y_j = \sum_{k=1}^N W_{jk} x_k$ and we have set $W \to \gamma W$ to have a coupling constant γ to tune. Note that the excitatory and inhibitory variables don't have to be labeled and ordered as above and the weight matrix W just describes a general signed network. From Theorem S52, we directly obtain the optimal reduced dynamics

$$\dot{X}_\mu = \sum_{\nu=1}^n \mathcal{D}_{\mu\nu} X_\nu + \sum_{j=1}^N M_{\mu j}(1 - a\textstyle\sum_{\nu=1}^n M^+_{j\nu} X_\nu)\, \mathcal{S}\left[b(\gamma \textstyle\sum_{\nu=1}^n \mathscr{W}_{j\nu} X_\nu - c)\right], \tag{S189}$$

where $\mathcal{D}_{\mu\nu} = -\sum_{j=1}^N M_{\mu j} d_j M^+_{j\nu}$ and $\mathscr{W}_{j\nu} = \sum_{k=1}^N W_{jk} M^+_{k\nu}$.

Under the form $Mf(M^+X)$ or, elements by elements, $\sum_{i=1}^N M_{\mu i} f_i(\sum_{\nu=1}^n M^+_{i\nu} X_\nu)$, there is still an explicit dependence of the vector field over N. Yet, we can sometimes eliminate this dependence by simplifying $Mf(M^+X)$ under certain properties of f which reveals something special about the resulting interaction between the observables.

C. Emergence of higher-order interactions

The critical role of higher-order interactions in complex systems is now increasingly recognized [187–191] and in this section, we aim at clarifying their origin by demonstrating the profound interplay between the description dimension of a system and the possibility of having higher-order interactions. When reducing the dimension of a dynamical system on a network, it is not always clear what to expect about the structure of the reduced dynamical system [see Fig. 3 in the paper]. We demonstrate that the structure that emerges from the dimension reduction in Theorem S52 generally yields higher-order interactions between the observables. For that, we first introduce some assumptions.

Assumptions S55.

(1) The N-dimensional dynamics on a network of weight matrix W is

$$\dot{x}_i = h_i(x_i, y_i), \quad i \in \{1, ..., N\}, \tag{S190}$$

where, for all i, $x_i : t \mapsto \mathbb{R}^N$, $y_i = \sum_{j=1}^N W_{ij} x_j$, and $h_i : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is an analytic function.

(2) The n-dimensional reduced dynamics is the least-square optimal dynamics [Theorem S52]

$$\dot{X}_\mu = \sum_{i=1}^N M_{\mu i} h_i(\tilde{x}_i, \tilde{y}_i), \quad \mu \in \{1, ..., n\}, \tag{S191}$$

where $X = Mx$ with any real reduction matrix M, $\tilde{x} = M^+X$, and $\tilde{y} = WM^+X$.

Condition (1) of Assumptions S55 might look restrictive because of the dependence over the linear function $x \mapsto Wx$. Yet, a considerable amount of complex system models satisfy condition (1) as shown in the following examples (from the power series in x_i, y_i of their analytic vector field, it is possible to classify the dynamics on networks of the next examples).

Example S56 (Epidemiology)**.** In the Susceptible-Infected-Susceptible (SIS) dynamics, an infected individual i (e.g., from a virus or disinformation) transmits its infection at a rate γ and recovers with rate d_i. In its exact form, the SIS dynamics is a homogeneous Markovian jump process and is described by master equations (forward Kolmogorov equations) [165, 192, 193]. Yet, since there are 2^N equations in this complete description and N is generally large, the typical approach is to consider some approximations of the process [90, 165, 193]. By neglecting the dynamical correlations between the states of the neighbors [194, Sec. 2.3.1], the quenched mean-field (QMF) approach [90] yields the deterministic system of equations

$$\dot{x}_i = -d_i x_i + \gamma(1 - x_i)\, y_i, \qquad i \in \{1, ..., N\}, \tag{S192}$$

called the QMF SIS model, where x_i is the probability for the vertex i to be infected. In Fig. 4, we use the latter dynamics as a simple introductory example. More generally, quenched mean-field approximations of many binary stochastic processes, such as the SIS dynamics above, the Cowan dynamics [185, 195], and the Glauber dynamics [196, 197] have the general form

$$\dot{x}_i = (1 - x_i)\,\alpha(k_i - y_i,\, y_i) + x_i\,\beta(k_i - y_i,\, y_i), \qquad i \in \{1, ..., N\}\,, \tag{S193}$$

where x_i is the probability for vertex i to be active, $k_i = \sum_{i=1}^{N} W_{ij}$ is the in-degree of vertex i, α (resp. β) is some analytic activation (resp. deactivation) probability function $\mathbb{R} \times \mathbb{R} \to [0, 1]$.

Example S57 (Neuroscience)**.** The Wilson-Cowan dynamics in Example S54 satisfies condition (1) of Assumption S55. Another popular model of neuronal activity, the threshold-linear model [198, 199], is defined by the equations

$$\dot{x}_i = -x_i + \left[\sum_{j=1}^{N} W_{ij} x_j + b_i\right]_+, \qquad i \in \{1, ..., N\}, \tag{S194}$$

where $y \mapsto [y]_+ = \max\{0, y\}$ is the standard rectifier or ReLU function. To meet condition (1), the latter must be replaced by an analytic approximation, such as the softplus function $y \mapsto \ln(1 + e^{ky})/k$ for some $k > 0$.

Example S58 (Population dynamics)**.** Population dynamics are widely used in science from ecology [200] and game theory [187] to chemistry (e.g., kinetic equations) [201] and physics (e.g., lasers) [202]. The generalized Lotka-Volterra model [203, 204] is a very typical population dynamics with the form

$$\dot{x}_i = -dx_i + \gamma x_i y_i\,. \tag{S195}$$

Refined models such as

$$\dot{x}_i = -dx_i - sx_i^2 + \gamma \frac{x_i y_i}{\alpha + y_i} \tag{S196}$$

in Ref. [200] or the microbial population dynamics [205]

$$\dot{x}_i = a - d\,x_i + bx_i^2 - cx_i^3 + \gamma\,x_i\,y_i. \tag{S197}$$

have also been used to incorporate more realistic effects, like the Allee effect in which a population exhibits negative growth for low abundances [206, 207]. In the latter dynamics, which is used in the paper, the correspondences with the parameters of Ref. [205] are $a = F$, $b = B(1 + K/C)$, $c = B/C$, and $d = BK$ where F is the migration rate, B is the logistic growth rate, C is the carrying capacity, and K is the Allee effect strength. In Extended Data Table 1, we consider that the parameter d can vary for each vertex only for the sake of coherence with the other dynamics.

Example S59 (Oscillators)**.** The Kuramoto-Sakaguchi dynamics [208, 209] is a canonical model for a large class of oscillatory systems [155, 210] and finds many applications, e.g., for Josephson junctions [211], nanoelectromechanical oscillators [212], and neuroscience [213]. The dynamics of the phase oscillators with a phase lag α is such that

$$\dot{\theta}_j = \omega_j + \gamma \sum_{k=1}^{N} W_{jk}\,\sin(\theta_k - \theta_j + \alpha), \tag{S198}$$

where $\theta_j(t)$ is the position of the j-th oscillator at time t, ω_j is the j-th natural frequency, and γ is the coupling constant. By setting $z_j = e^{i\theta_j}$ [154], the Kuramoto-Sakaguchi model becomes

$$\dot{z}_j = i\omega_j z_j + \gamma\,e^{-i\alpha}\,y_j - \gamma\,e^{i\alpha}\,z_j^2\,\bar{y}_j, \tag{S199}$$

where $y_j = \sum_{k=1}^{N} W_{jk} z_k$ and $\bar{\ }$ denotes complex conjugation. Note that the Winfree model [214] and the theta model [215] on networks [154] also satisfy the condition (1) of Assumption S55.

Example S60 (Machine learning)**.** The universal approximation theorem of Funahashi and Nakamura [216, Theorem 1] guarantees that a solution to a general dynamical system is approximately given, up to the desired accuracy, by a solution of a continuous-time recurrent neural network [216, 217]

$$\dot{x}_i = -\frac{1}{\tau_i} x_i + \sum_{j=1}^{N} W_{ij} \mathcal{S}(x_j) + I_i, \tag{S200}$$

where x_i is the trajectory of the i-th neuron, τ_i is the time-scale of neuron i, $\mathcal{S}$ is the sigmoid (logistic) function, W_{ij} is the element (i, j) of the $N \times N$ weight matrix W, and I_i is the input current applied on neuron i. Equation (S200) is also called or similar, up to some variations in its form, Cohen-Grossberg model [218, 219], Hopfield model [220], activation dynamics [221], continuous rate RNN [222, 223], or reservoir computers. This recurrent neural network does not directly have the form to satisfy the condition (1) of Assumption S55, but from Ref. [224], we know that there is a class of (continuous-time) recurrent neural networks with the form

$$\dot{x}_i = -d_i x_i + \tanh(\gamma y_i + c_i), \tag{S201}$$

$d_1, ..., d_N$ are real constants, $c_i : t \mapsto c_i(t) \in \mathbb{R}$ is the i-th current, that is a universal approximator and satisfy the condition (1).

Following these considerations, we introduce a general proposition about the emergence of higher-order interactions when reducing the dimension of a dynamical system on network using Theorem S52.

Proposition S61. *If the conditions of Assumptions S55 hold, the reduced dynamics can be expressed in terms of higher-order interactions between the observables as*

$$\dot{X}_\mu = \mathcal{C}_\mu + \sum_{d_x=1}^{\infty} \sum_{\boldsymbol{\alpha}\in\mathbb{Z}_+^n} \mathcal{D}_{\mu\boldsymbol{\alpha}}^{(d_x+1)} X_{\boldsymbol{\alpha}} + \sum_{d_y=1}^{\infty} \sum_{\boldsymbol{\beta}\in\mathbb{Z}_+^n} \mathcal{W}_{\mu\boldsymbol{\beta}}^{(d_y+1)} X_{\boldsymbol{\beta}} + \sum_{d_x,d_y=1}^{\infty} \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{Z}_+^n} \mathcal{T}_{\mu\boldsymbol{\alpha}\boldsymbol{\beta}}^{(d_x+d_y+1)} X_{\boldsymbol{\alpha}\boldsymbol{\beta}}, \tag{S202}$$

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

$$\mathcal{D}_{\mu\boldsymbol{\alpha}}^{(d_x+1)} = \sum_{i=1}^{N} c_{id_x0} M_{\mu i} M_{i\alpha_1}^+ \dots M_{i\alpha_{d_x}}^+, \tag{S203}$$

$$\mathcal{W}_{\mu\boldsymbol{\beta}}^{(d_y+1)} = \sum_{i,j_1,\dots,j_{d_y}=1}^{N} c_{i0d_y} M_{\mu i} W_{ij_1} ... W_{ij_{d_y}} M_{j_1\beta_1}^+ \dots M_{j_{d_y}\beta_{d_y}}^+, \tag{S204}$$

$$\mathcal{T}_{\mu\boldsymbol{\alpha}\boldsymbol{\beta}}^{(d_x+d_y+1)} = \sum_{i,j_1,\dots,j_{d_y}=1}^{N} c_{id_xd_y} M_{\mu i} M_{i\alpha_1}^+ \dots M_{i\alpha_{d_x}}^+ W_{ij_1} ... W_{ij_{d_y}} M_{j_1\beta_1}^+ \dots M_{j_{d_y}\beta_{d_y}}^+, \tag{S205}$$

for some real coefficients $c_{id_xd_y}$ with $i \in \{1, ..., N\}$ and $d_x, d_y \in \mathbb{Z}_+$.

Proof. By definition of an analytic function, there is a convergent power series describing the vector field of the complete dynamics, i.e.,

$$h_i(x_i, y_i) = \sum_{d_x=0}^{\infty} \sum_{d_y=0}^{\infty} c_{id_xd_y} x_i^{d_x} y_i^{d_y}, \qquad i \in \{1, ..., N\}, \tag{S206}$$

where we have chosen to express the power series around $x_i = y_i = 0$ without loss of generality. The reduced dynamics is therefore

$$\dot{X}_\mu = \sum_{d_x,d_y=0}^{\infty} \sum_{i=1}^{N} c_{id_xd_y} M_{\mu i} \left(\sum_{\alpha=1}^{n} M_{i\alpha}^+ X_\alpha \right)^{d_x} \left(\sum_{j=1}^{N} \sum_{\beta=1}^{n} W_{ij} M_{j\beta}^+ X_\beta \right)^{d_y}.$$

The sum can be separated as

$$\begin{aligned}\dot{X}_\mu = \mathcal{C}_\mu &+ \sum_{d_x=1}^{\infty} \sum_{i=1}^{N} c_{id_x0} M_{\mu i} \left(\sum_{\alpha=1}^{n} M_{i\alpha}^+ X_\alpha \right)^{d_x} + \sum_{d_y=1}^{\infty} \sum_{i=1}^{N} c_{i0d_y} M_{\mu i} \left(\sum_{j=1}^{N} \sum_{\beta=1}^{n} W_{ij} M_{j\beta}^+ X_\beta \right)^{d_y} \\ &+ \sum_{d_x,d_y=1}^{\infty} \sum_{i=1}^{N} c_{id_xd_y} M_{\mu i} \left(\sum_{\alpha=1}^{n} M_{i\alpha}^+ X_\alpha \right)^{d_x} \left(\sum_{j=1}^{N} \sum_{\beta=1}^{n} W_{ij} M_{j\beta}^+ X_\beta \right)^{d_y},\end{aligned}$$

where we have defined $\mathcal{C}_\mu = \sum_{i=1}^{N} M_{\mu i} c_{i00}$. Expanding the exponents and introducing the multi-indices directly provide the desired result. □

Remark S62.

1. As explained in Section 1.1 (p.3) of Ref. [225], the tensors above could be more precisely called hypermatrices.
2. For clarity, we specify the order of the tensor as an exponent in parentheses. In the paper and in Example S68, the order is clear from the indices and we thus avoid this notation for simplicity. Also, we let the indices differentiate the tensors, e.g., $\mathcal{T}_{1(2,3)(4)}$ ($d_x = 2$, $d_y = 1$) and $\mathcal{T}_{1(2)(3,4)}$ ($d_x = 1$, $d_y = 2$) are elements of two different tensors. Finally, when it's clear in the context, if a multi-index is a singleton, than we remove the parentheses, e.g., $\mathcal{T}_{\mu(\nu)(\kappa)}$ becomes $\mathcal{T}_{\mu\nu\kappa}$.
3. The coefficients $c_{id_xd_y}$ can be chosen as the ones of the Taylor series of h_i for all i.

4. For the sake of simplicity, let us consider the case where $c_{i01} = 1$ for all i. We observe that $\mathcal{W}^{(2)} = MWM^+$ appears in the reduced dynamics, which can be viewed as the reduced weight matrix. From Theorem S50, it is also the unique solution to the compatibility equation [154] $\mathcal{W}M = MW$ when $\operatorname{rank} M = n$ and it is the least-square optimal solution to the problem $\|\mathcal{W}M - MW\|_F^2$ with $\mathcal{W}$ as the optimization variable. Remember from Ref. [154] that solving the compatibility equation is necessary to cancel the first-order errors in DART or less generally, to close the reduced dynamics of any linear dynamics $\dot{x} = Wx$. Indeed, for $X = Mx$, $\dot{X} = MWx = \mathcal{W}Mx = \mathcal{W}X$ where one can reasonably choose $\mathcal{W} = \mathcal{W}^{(2)}$ as explained before.
5. If there was already higher-order interactions in the complete dynamics, the least-square optimal reduced dynamics would have new higher-order interactions that depends on the original ones, the parameters of the dynamics, and the reduction matrix.
6. The latter proposition can easily be extended to complex variables. First assume that the complex dynamics has the form $\dot{x}_i = r_i(x_i, y_i, \bar{x}_i, \bar{y}_i)$, where $\bar{\ }$ is complex conjugation and $r_i : \mathbb{C}^4 \mapsto \mathbb{C}$ is a holomorphic function (and thus analytic):
$$r_i(x_i, y_i, \bar{x}_i, \bar{y}_i) = \sum_{d_x=0}^{\infty} \sum_{d_y=0}^{\infty} \sum_{\bar{d}_x=0}^{\infty} \sum_{\bar{d}_y=0}^{\infty} c_{id_xd_y\bar{d}_x\bar{d}_y} x_i^{d_x} y_i^{d_y} \bar{x}_i^{\bar{d}_x} \bar{y}_i^{\bar{d}_y}, \quad i \in \{1, ..., N\}. \tag{S207}$$
The rest of the proof is similar to its real counterpart. This is especially interesting for phase dynamics such as the Kuramoto model (see Example S68).
7. This is not the only dimension reduction that yields higher-order interactions. We did not realize it clearly at the moment of writing Ref. [154], but DART also yields higher-order interactions, which can be explicitly seen in Eqs. (28-30). However, these higher-order interactions could be avoided by noting that the phase dynamics have a vector field of the form $h_i(x_i, y_i)$. Indeed, using Taylor's theorem for both x and Wx, there is no compatibility equation for the degrees that appears to cancel the first-order terms and it ultimately removes the higher-order contributions with $\mathcal{K}$ in Eqs. (28-30). In general, a dimension reduction method where the original vector field is evaluated at a function of the original variables is susceptible to yield higher-order interactions.

In the last proposition, the graph with N vertices of the complete dynamics (and its parameters encoded by all the coefficients $c_{id_xd_y}$) is thus replaced by a hypergraph $\mathcal{H}$ [225–227] with n vertices [see Fig. 3d of the paper], defined from the tensors $\mathcal{D}^{(d_x+1)}$, $\mathcal{W}^{(d_y+1)}$, and $\mathcal{T}^{(d_x+d_y+1)}$, in the reduced dynamics. Below, we define more precisely the notion of directed, weighted, and signed hypergraphs.

Definition S63. A hypergraph is a triple $\mathcal{H} = (\Upsilon, \Xi, \Omega)$, where
- $\Upsilon = \{1, ..., n\}$ is the set of vertices;
- Ξ is a set of *hyperarcs* (or directed hyperedges) defined as an ordered pair $E = (H, T)$, where H is the *head* of the hyperarc (a n_H-tuple with elements in Υ), T is the *tail* of the hyperarc (a n_T-tuple with elements in Υ), and $2 \leq n_H + n_T \leq n$ with $n_H, n_T \geq 1$. For $n_H = 1$ and $n_T = 1$, the hyperarc is a directed edge. If $n_H = 1$ and $n_T > 1$, it is a backward hyperarc and if $n_H > 1$ and $n_T = 1$, it is a forward hyperarc;
- Ω is a function that assigns a real value to the hyperarcs.

Remark S64.
- The latter definition is a generalization of hypergraphs [226] and of directed hypergraphs as defined in Ref. [227], where the head and the tails of the hyperarcs are sets instead of tuples.
- For the weight matrix with elements W_{ij}, we use the convention that the edge (or arc) (i, j) is directed from j to i. For consistency, in the definition above, we use the convention that the hyperarc (H, T) (instead of (T, H) as in Ref. [227]) is directed from the tail T to the head H. As a consequence, in the tensor notation $\mathcal{T}^{(d_x+d_y+1)}_{\mu\boldsymbol{\alpha}\boldsymbol{\beta}}$, the index μ and the multi-index $\boldsymbol{\alpha}$ are part of the head while $\boldsymbol{\beta}$ is part of the tail of the hyperarc. Thus, $\mathcal{T}_{1(2)(3)}$ $(n_T = 1)$ is a forward hyperarc $((1, 2), (3))$ while $\mathcal{T}_{1()(2,3)}$ $(n_H = 1)$ is a backward hyperarc $(1, (2, 3))$. Note that the tensor $\mathcal{W}^{(d_y+1)}$ with elements in Eq. (S204) always form backward hyperarcs (from $\boldsymbol{\beta}$ to μ) since $d_x = 0$, while the tensor with elements in Eq. (S205) can be any type of hyperarc (with μ always belonging to the head). In the example of the paper for the epidemiological dynamics, Eq. (7) is a forward hyperarc (from κ to $\mu\nu$).

We now derive two key consequences of Proposition S61. First, Proposition S61 shows that there can be an infinite number of higher-order interactions in the reduced dynamics. Yet, for a special family of vector fields, we prove that there is a finite number of them which are related to the nonlinearity of the original dynamics.

Corollary S65. *If $h_i(x_i, y_i)$ is a polynomial of total degree δ in x_i and y_i for all $i \in \{1, ..., N\}$ and condition (2) of Assumptions S55 holds, then the reduced dynamics has a polynomial vector field of total degree δ with interactions of maximal order $\delta + 1$.*

Proof. Since any polynomial is analytic, condition (1) of Assumptions S55 is satisfied. Then, by Proposition S61, the reduced dynamics is given by Eqs. (S202-S205). In the following, the conclusions are valid for all $i \in \{1, ..., N\}$. Let

$\mathscr{I}_i = \{c_{id_xd_y}\}_{d_x,d_y=0}^{\infty}$ be the i-th (countable) infinite set of coefficients related to the i-th analytic function h_i. The fact that h_i is a polynomial implies that there is a finite subset of nonzero coefficients $\mathscr{F}_i \subset \mathscr{I}_i$ describing a polynomial vector field for the reduced dynamics. Consider any coefficient $c_{i'd'_xd'_y} \in \mathscr{F}_i$ such that $d'_x + d'_y = \delta$, the total degree of the polynomial h_i. Then, at least one of the tensors $\mathcal{D}^{(d'_x+1)}$, $\mathcal{W}^{(d'_y+1)}$, $\mathcal{T}^{(d'_x+d'_y+1)}$, with elements in Eqs. (S203-S205), have the highest possible order $\delta+1$. Moreover, there will be at least one monomial term $X_{\alpha_1}...X_{\alpha_{d'_x}}$, $X_{\beta_1}...X_{\beta_{d'_y}}$, or $X_{\alpha_1}...X_{\alpha_{d'_x}}X_{\beta_1}...X_{\beta_{d'_y}}$ in Eq. (S202) that is of maximal degree δ, which means that reduced dynamics has a polynomial vector field of total degree δ. □

Second, the tensors describe Proposition S61 strongly depends on the reduction matrix M, or in other words, the reduction matrix will play a role on the form of the higher-order interactions. One can therefore ask if one can choose M in such a way that there are only pairwise interactions in the reduced dynamics. In the next corollary, we provide sufficient conditions to have pairwise interactions in the least-square reduced dynamics.

Corollary S66. *Let $s : \mathcal{V} \to \Upsilon$ be a surjection where $\mathcal{V} = \{1, ..., N\}$ and $\Upsilon \in \{1, ..., n\}$ are the vertex sets of the complete and reduced system respectively. If Assumptions S55 hold, the reduction matrix M has elements $M_{\mu i} = m_{\mu i}\delta_{\mu\, s(i)}$ with $m_{\mu i} \in \mathbb{R}$ for all μ, i, and h_i linearly depends on y_i for all i, then there are solely pairwise interactions in the reduced system. The result doesn't hold in general for nonlinear dependencies of h_i over y_i.*

Proof. For such reduction matrix, the elements of its Moore-Penrose pseudoinverse are, for all $\mu \in \Upsilon$ and $i \in \mathcal{V}$, $M^+_{i\mu} = m_{\mu i}\delta_{s(i)\mu}/q_\mu$, where $q_\mu = \sum_{i=1}^N m_{\mu i}^2 \delta_{s(i)\mu}$. Substituting M and M^+ in Eqs. (S203-S205) yields

$$\mathcal{D}^{(d_x+1)}_{\mu\boldsymbol{\alpha}} = \frac{1}{q_{\boldsymbol{\alpha}}} \sum_{i=1}^{N} c_{id_x0} m_{\mu\boldsymbol{\alpha} i}\, \delta_{\mu\, s(i)} \delta_{s(i)\,\alpha_1} \dots \delta_{s(i)\,\alpha_{d_x}},$$

$$\mathcal{W}^{(d_y+1)}_{\mu\boldsymbol{\beta}} = \frac{1}{q_{\boldsymbol{\beta}}} \sum_{i,j_1,\dots,j_{d_y}=1}^{N} c_{i0d_y} m_{\mu i}\, \delta_{\mu\, s(i)} W_{ij_1}...W_{ij_{d_y}} \delta_{s(j_1)\,\beta_1} \dots \delta_{s(j_{d_y})\,\beta_{d_y}},$$

$$\mathcal{T}^{(d_x+d_y+1)}_{\mu\boldsymbol{\alpha\beta}} = \frac{1}{q_{\boldsymbol{\alpha\beta}}} \sum_{i,j_1,\dots,j_{d_y}=1}^{N} c_{id_xd_y} m_{\mu\boldsymbol{\alpha} i}\, \delta_{\mu\, s(i)} \delta_{s(i)\,\alpha_1} \dots \delta_{s(i)\,\alpha_{d_x}} W_{ij_1}...W_{ij_{d_y}} m_{\beta_1 j_1}...m_{\beta_{d_y} j_{d_y}} \delta_{s(j_1)\,\beta_1} \dots \delta_{s(j_{d_y})\,\beta_{d_y}},$$

where $q_{\boldsymbol{\gamma}} = q_{\gamma_1}...q_{\gamma_d}$ and $m_{\mu\boldsymbol{\alpha} i} = m_{\mu i} m_{\alpha_1 i}...m_{\alpha_d i}$. For $\mathcal{D}^{(d_x+1)}$ and any dependence of h_i over y_i, it is straightforward to observe that the only nonzero elements are such that $\mu = \alpha_1 = ... = \alpha_{d_x}$. The tensor can therefore be mapped to a $n \times n$ diagonal matrix. Henceforth, we only consider $\mathcal{W}^{(d_y+1)}$ and $\mathcal{T}^{(d_x+d_y+1)}$.

The fact that $h_i(x_i, y_i)$ linearly depends on y_i for all i is equivalent to setting $d_y = 1$ in its power series in Eq. (S206), i.e.,

$$h_i(x_i, y_i) = \sum_{d_x=0}^{\infty} c_{id_x1} x_i^{d_x} y_i, \qquad i \in \{1, ..., N\}.$$

Proposition S61 thus implies that

$$\mathcal{W}^{(2)}_{\mu\beta} = \frac{1}{q_\beta} \sum_{i,j=1}^{N} c_{i01} m_{\mu i} m_{\beta j} \delta_{\mu\, s(i)} W_{ij} \delta_{s(j)\,\beta} \;\text{ and }\; \mathcal{T}^{(d_x+2)}_{\mu\boldsymbol{\alpha}\beta} = \frac{1}{q_{\boldsymbol{\alpha}\beta}} \sum_{i,j=1}^{N} c_{id_x1} m_{\mu\boldsymbol{\alpha} i} \delta_{\mu\, s(i)} \delta_{s(i)\,\alpha_1} \dots \delta_{s(i)\,\alpha_{d_x}} W_{ij} m_{\beta j}\, \delta_{s(j)\,\beta}.$$

Clearly, $\mathcal{W}^{(2)}$ is a matrix and the nonzero elements of $\mathcal{T}^{(d_x+2)}$ are for $\mu = \alpha_1 = ... = \alpha_{d_x}$ (there are at most n^2 of them), which means that it can be mapped to a $n \times n$ matrix. Hence, there are solely pairwise interactions in the least-square reduced dynamics.

If $d_y > 1$ (i.e., for a nonlinear dependency of h_i over y_i), a simple example suffices to prove the last statement. Let $d_y = 2$, $\mathcal{V} = \{1, 2, 3, 4, 5\}$, $\Upsilon = \{1, 2, 3\}$, and $s(1) = 1$, $s(2) = 1$, $s(3) = 2$, $s(4) = 3$, $s(5) = 3$. Moreover, consider that W_{52}, W_{53}, m_{12}, m_{23}, m_{35}, c_{502} are not equal to zero. Then, Proposition S61 gives

$$\mathcal{W}^{(3)}_{\mu(\beta_1,\beta_2)} = \frac{1}{q_{\mu\beta_1\beta_2}} \sum_{i=1}^{N} c_{i02} m_{\mu i}\, \delta_{\mu\, s(i)} \left(\sum_{j,k=1}^{N} m_{\beta_1 j} m_{\beta_2 k} W_{ij} W_{ik} \delta_{s(j)\,\beta_1} \delta_{s(k)\,\beta_2} \right).$$

It only remains to prove that there can be nonzero elements for $\beta_1 \neq \beta_2$. For $\beta_1 = 1$, $\beta_2 = 2$, $j = 2$, and $k = 3$ in the parentheses of the last equation, there is a term $m_{12}m_{23}W_{i2}W_{i3}\delta_{s(2)\,1}\delta_{s(3)\,2} = m_{12}m_{23}W_{i2}W_{i3}$. Considering the whole equation for $i = 5$ and $\mu = 3$, there is a nonzero term $c_{502}m_{35}m_{12}m_{23}W_{52}W_{53}/q_{312}$. Hence, in this example, $\mathcal{W}_{3(1,2)} \neq 0$ despite the fact that the observables are defined on disjoint sets of vertices. □

Remark S67. If $n = N$, the higher-order interactions between the observables does not necessarily disappear because of the linear transformation done by M on x. Obviously, if $M = I$, $X = x$ and $\dot{X}_\mu = \dot{x}_i = h_i(x_i, y_i)$ for all i and there are no higher order interactions. However, the vector field $M \circ h \circ M^+$ will generally contain higher-order interactions. But of course, if M has full rank N, it is invertible and one can transform back the dynamics of the observable X (with higher-order interactions) to the dynamics in x (without higher-order interactions), since $x = M^{-1}X$.

Let's now provide the details about the examples presented in Extended Data Table 1.

Example S68 (Emergence of higher-order interactions in typical models)**.** Proposition S61 and Corollary S65 imply the following results in different fields of application.

1. QMF SIS dynamics [Eq. (S192) in Example S56]:

$$\dot{X}_\mu = \sum_{\nu=1}^{n} (\mathcal{D}_{\mu\nu} + \mathcal{W}_{\mu\nu}) X_\nu + \sum_{\nu,\kappa=1}^{n} \mathcal{T}_{\mu\nu\kappa} X_\nu X_\kappa, \tag{S208}$$

   where $\mathcal{D} = -MDM^+$ with $D = \mathrm{diag}(d_1, ..., d_N)$, $\mathcal{W} = \gamma MWM^+$, and

$$\mathcal{T}_{\mu\nu\kappa} = -\gamma \sum_{i,j=1}^{N} M_{\mu i} M^+_{i\nu} W_{ij} M^+_{j\kappa},$$

   with $\boldsymbol{\alpha} = (\nu) = \nu$ and $\boldsymbol{\beta} = (\kappa) = \kappa$. Interestingly, for $n = 1$, one can find the exact solution to Eq. (S192) since it is a Bernouilli differential equation.

2. Microbial population dynamics [Eq. (S197) in Example S58]:

$$\dot{X}_\mu = \mathcal{C}_\mu + \sum_{\nu=1}^{n} \mathcal{D}_{\mu\nu} X_\nu + \sum_{\nu,\kappa=1}^{n} (\mathcal{D}_{\mu(\nu,\kappa)} + \mathcal{T}_{\mu\nu\kappa}) X_\nu X_\kappa + \sum_{\nu,\kappa,\tau=1}^{n} \mathcal{D}_{\mu(\nu,\kappa,\tau)} X_\nu X_\kappa X_\tau \tag{S209}$$

   where $\mathcal{D} = -dMM^+$ and

$$\mathcal{D}_{\mu(\nu,\kappa)} = b \sum_{i=1}^{N} M_{\mu i} M^+_{i\nu} M^+_{i\kappa}\,, \quad \mathcal{T}_{\mu\nu\kappa} = \gamma \sum_{i,j=1}^{N} M_{\mu i} M^+_{i\nu} W_{ij} M^+_{j\kappa}\,, \quad \mathcal{D}_{\mu(\nu,\kappa,\tau)} = -c \sum_{i=1}^{N} M_{\mu i} M^+_{i\nu} M^+_{i\kappa} M^+_{i\tau}\,.$$

3. Kuramoto-Sakaguchi dynamics [Eq. (S199) in Example S59]:

$$\dot{X}_\mu = \sum_{\nu=1}^{n} (\mathcal{D}_{\mu\nu} + \mathcal{W}_{\mu\nu}) X_\nu + \sum_{\nu,\kappa,\tau=1}^{n} \mathcal{T}_{\mu(\nu,\kappa)\tau} X_\nu X_\kappa \bar{X}_\tau \tag{S210}$$

   where $\mathcal{D} = iMDM^+$ with $D = \mathrm{diag}(\omega_1, ..., \omega_N)$, $\mathcal{W} = \gamma e^{-i\alpha} MWM^+$, and

$$\mathcal{T}_{\mu(\nu,\kappa)\tau} = -\gamma e^{i\alpha} \sum_{j,k=1}^{N} M_{\mu j} M^+_{j\nu} M^+_{j\kappa} W_{jk} M^+_{k\tau},$$

   with $\boldsymbol{\alpha} = (\nu, \kappa)$, $\boldsymbol{\beta} = (\tau) = \tau$. In this case, the reduced variables X_1,...,X_n and the involved tensors are complex.

Remark S69.

- We found that there can be computational benefits to write the vector fields in terms of tensors (subsection III G).
- The fact that the least-square optimal reduced vector field contains higher-order interactions raises the problem of getting mathematical insights from dynamics on hypergraphs, which recalls again the pertinence of this field in the study of complex systems. Fortunately, many recent papers address the problem, such as Ref. [228] or Ref. [229]. See Ref. [190] for more references.

In phase reduction techniques [210], dN-dimensional weakly coupled limit-cycle oscillators dynamics, where each of the N oscillators is described by d variables, are reduced to a N-dimensional dynamics of their phase. It is known that these phase reductions lead to higher-order interactions between the phases [212, 230, 231] or, in other words, between microscopic observables (i.e., there is a phase for each oscillator, considered as the microscopic level, except in Ref. [232, Fifth section]). In contrast, the higher-order interactions that we observe emerge from a large variety of dynamical systems and they are between observables that can cover different scales, which strongly depends over the choice reduction matrix. The generality of our results thus suggests that the emergence could be quite ubiquitous.

D. Upper bound on the alignment error and exact dimension reduction

In this subsection, we evaluate the impact of choosing the least-square optimal vector field in Theorem S52 on the alignment error $\mathcal{E}(x)$ in $\mathbb{R}^n$. In particular, we will see that obtaining an upper bound on $\mathcal{E}(x)$ is useful to find a reasonable choice of reduction matrix M. More importantly, to determine more quantitatively the repercussions of the low-rank hypothesis on the dynamics, we aim at estimating the error caused by the optimal reduced dynamics as a function of n. Let us start by listing the assumptions that will be made throughout this subsection.

Assumptions S70.

(1) The N-dimensional complete dynamics on a network defined by the real $N \times N$ weight matrix W is

$$\dot{x} = g(x, y) = \begin{pmatrix} g_1(x_1, ..., x_N, y_1, ..., y_N) \\ \vdots \\ g_N(x_1, ..., x_N, y_1, ..., y_N) \end{pmatrix}, \tag{S211}$$

where $x : t \mapsto \mathbb{R}^N$, $y = Wx$, and $g : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ is a continuously differentiable function.

(2) The n-dimensional reduced dynamics ($n < N$) is the least-square optimal dynamics of Theorem S52, i.e.,

$$\dot{X} = Mg(M^+X, WM^+X)\,. \tag{S212}$$

(3) The reduction matrix M is the truncated left singular vector matrix $V_n^\top$ of W.

Note the first assumption is less restrictive than the first one of the Assumptions S55. We chose $n < N$ to ensure dimension reduction and also, because it is obvious to show that we can have a zero alignment error when $n = N = \operatorname{rank} M$. In this case, the "reduced" dynamics is not reduced anymore, but it is still a linear transformation of the complete dynamics.

Lemma S71. *If conditions (1) and (2) in Assumptions S70 hold with $n = N = \operatorname{rank} M$, the alignment error is 0.*

Proof. If M has rank N, the pseudoinverse of M is its inverse and the related projector is $P = M^+M = M^{-1}M = I$. Hence, the alignment error in $\mathbb{R}^n$ is obviously zero: $\mathcal{E}(x) = \|M[g(x, Wx) - g(Px, WPx)]\|/\sqrt{n} = 0$. ☐

Let us now turn to one of the important results of the paper. The next theorem demonstrates that the alignment error between a high-dimensional vector field depending on a network and its optimally reduced version is intrinsically related to the network's singular value profile: when the singular values σ_n decrease rapidly with n, so does the alignment error. Therefore, a low-rank hypothesis induces a low-dimension hypothesis for dynamical systems.

Theorem S72. *If all conditions of Assumptions S70 hold, the alignment error in $\mathbb{R}^n$ at $x \in \mathbb{R}^N$ is upper-bounded as*

$$\mathcal{E}(x) \leq \frac{1}{\sqrt{n}}\Big[\|V_n^\top J_x(x', y')(I - V_nV_n^\top)x\| + \sigma_{n+1}\|V_n^\top J_y(x', y')\|_2\|x\|\Big], \tag{S213}$$

where $y' = Wx'$ with x' being some point between x and $V_nV_n^\top x$, σ_i is the i-th singular value of W, and $J_x(x', y')$, $J_y(x', y')$ are the Jacobian matrices of g with derivatives according to the vectors x and y respectively. Moreover, for any x not at the origin of $\mathbb{R}^N$, the following upper bound on the relative alignment error holds:

$$\frac{\mathcal{E}(x)}{\|x\|} \leq \frac{1}{\sqrt{n}}\Big[\alpha(x', y') + \sigma_{n+1}\beta(x', y')\Big]\,, \tag{S214}$$

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

Proof. From the definition of the alignment error and the first two conditions in Assumptions S70, we have

$$\mathcal{E}(x) = \frac{1}{\sqrt{n}}\|M[g(x, y) - g(\tilde{x}, \tilde{y})]\|,$$

where $y = Wx$, $\tilde{x} = Px$, and $\tilde{y} = WPx$ with $P = M^+M$. Let's define the function

$$u(x) = g(x, \ell(x)), \tag{S215}$$

with the linear function $\ell(x) = Wx$. Since g is a continuously differentiable function, u is also continuously differentiable and Taylor's theorem with 0-th order Lagrange remainder guarantees that

$$u(x) = u(\tilde{x}) + Du(x')(x - \tilde{x}) \tag{S216}$$

for some x' between x and Px and where $Du(x')$ is the total derivative of u. From Eq. (S215) and the chain rule for the total derivative, we have (abusing the matrix notation)

$$Du(x') = Dg(x', y') = \frac{\partial g}{\partial x}(x', y') + \frac{\partial g}{\partial y}(x', y')\frac{\partial \ell}{\partial x}(x') = J_x(x', y') + J_y(x', y')W, \tag{S217}$$

where $y' = \ell(x') = Wx'$, the elements of the Jacobian matrices $J_x(x', y'), J_y(x', y')$ are respectively

$$[J_x(x', y')]_{ij} = \frac{\partial g_i(x, y)}{\partial x_j}\bigg|_{(x,y)=(x',y')}, \quad [J_y(x', y')]_{ij} = \frac{\partial g_i(x, y)}{\partial y_j}\bigg|_{(x,y)=(x',y')},$$

and we have used the fact that W is the Jacobian matrix of ℓ. The Taylor expansion (S216) of u with N variables $(x_i)_{i=1}^N$ for some x' therefore implies a Taylor expansion for g with $2N$ variables $(x_i, y_i)_{i=1}^N$ for some x' with $y' = Wx'$:

$$g(x, y) = g(\tilde{x}, \tilde{y}) + J_x(x', y')(x - \tilde{x}) + J_y(x', y')W(x - \tilde{x}). \tag{S218}$$

The alignment error becomes

$$\mathcal{E}(x) = \frac{1}{\sqrt{n}}\|M[J_x(x', y')(I - P)x + J_y(x', y')W(I - P)x]\|$$

and the triangle inequality gives

$$\mathcal{E}(x) \leq \frac{1}{\sqrt{n}}\big[\|MJ_x(x', y')(I - P)x\| + \|MJ_y(x', y')W(I - P)x\|\big].$$

Moreover, the induced spectral norm for the second term yields

$$\mathcal{E}(x) \leq \frac{1}{\sqrt{n}}\big[\|MJ_x(x', y')(I - P)x\| + \|MJ_y(x', y')W(I - P)\|_2\|x\|\big]$$

and the submultiplicativity of the spectral norm implies

$$\mathcal{E}(x) \leq \frac{1}{\sqrt{n}}\big[\|MJ_x(x', y')(I - P)x\| + \|W(I - P)\|_2\|MJ_y(x', y')\|_2\|x\|\big]. \tag{S219}$$

From condition (3) of Assumption S70, we have $M = V_n^\top$ which is, by Theorem S9, the optimal solution to the minimization of $\|W(I - P)\|_2$ with error σ_{n+1} (square root of the problem (P2) and the error for the spectral norm in Eq. (S11)). The second inequality is deduced as follows:

$$\begin{aligned}
\frac{\mathcal{E}(x)}{\|x\|} &\leq \frac{1}{\sqrt{n}}\Big[\|V_n^\top J_x(x', y')(I - V_nV_n^\top)\|_2 + \sigma_{n+1}\|V_n^\top J_y(x', y')\|_2\Big] && \text{(S220)}\\
&\leq \frac{1}{\sqrt{n}}\Big[\|V_n^\top J_x(x', y')\|_2 + \sigma_{n+1}\|V_n^\top J_y(x', y')\|_2\Big] && \text{(S221)}\\
&\leq \frac{1}{\sqrt{n}}\Big[\|J_x(x', y')\|_2 + \sigma_{n+1}\|J_y(x', y')\|_2\Big], && \text{(S222)}
\end{aligned}$$

where we have used successively the submultiplicativity of the spectral norm and identities $\|(I - V_nV_n^\top)\|_2 = 1$, $\|V_n^\top\|_2 = 1$. The desired upper bound is found upon noticing that $\|J_x(x', y')\|_2 = \sigma_1(J_x(x', y'))$ and $\|J_y(x', y')\|_2 = \sigma_1(J_y(x', y'))$. □

Remark S73.

- The dynamics used in the paper have the less general form (compared to condition (1) in Assumptions S70)

$$\dot{x}_i = h_i(x_i, y_i), \quad i \in \{1, ..., N\}, \tag{S223}$$

  where, for all i, $x_i : t \mapsto \mathbb{R}^N$, $y_i = \sum_{j=1}^N W_{ij}x_j$ and $h_i : \mathbb{R}^2 \mapsto \mathbb{R}$. This implies that for all dynamics considered in the paper, the Jacobian matrices $J_x(x', y')$ and $J_y(x', y')$ are diagonal.
- Even if the effective ranks of real networks are low compared to N, they are generally larger than one, meaning that σ_{n+1} is not negligible when $n = 1$. According to our analysis, we therefore do not expect one-dimensional reductions [207, 233, 234] to yield accurate results in general, which is consistent with numerical observations made in previous studies [154, 233, 235–237]. Some very simple synthetic networks, however, such as those generated by the Erdös-Rényi and Chung-Lu models, typically have a very small second singular value, suggesting that accurate one-dimensional reductions are possible for those cases.
- Using the induced spectral norm in the upper bound introduces a factor of about $\sqrt{N}/2$ when sampling x uniformly between 0 and 1. This is one of the main reasons why the bound is not always tight. But our focus is not on magnitude of the error or the tightness of the bound, but on the decrease of the error. The extra $\sqrt{N}$

is removed by considering the relative alignment error.

- The relative alignment error $\mathcal{E}(x)/\|x\|$ is upper-bounded by purely spectral factors, which can be classified into two types: (1) those related to the Jacobians and thus depending upon the dynamics, and (2) σ_{n+1} that only depends on the network. The second type is universal in the sense that it applies to all dynamics. Contrary to what is observed with σ_{n+1} in real networks, the factors $\alpha(x', y')$ and $\beta(x', y')$ do not necessarily decrease as n increases.
- The relative alignment error bound above can be improved, but it has a price. Indeed, the first steps of the theorem, the induced spectral norm and the submultiplicativity lead to $\mathcal{E}(x)/\|x\| \leq \|M\|_2 \|Du(x')(I-P)\|_2/\sqrt{n}$. From there, one could consider that M is dependent over x' and set $M := R_n^\top(x')$, the truncated right singular vector matrix of the Jacobian matrix $Du(x')$. Again, this choice minimizes $\|Du(x')(I-P)\|_2$ from Theorem S9 and one has the simple upper bound $\mathcal{E}(x)/\|x\| \leq \gamma_{n+1}(x')/\sqrt{n}$, where $\gamma_{n+1}(x')$ is the $(n+1)$-th singular value of $Du(x')$ depending on x'. The dependence of the reduction matrix over x' is, however, not desired since we want the reduced dynamics to be independent of the N-dimensional dynamics.

As a byproduct of the last theorem, the fact that the term $\|W(I - M^+M)\|_2$ appears in the upper bound in Eq. (S219) of the alignment error suggests a reasonable choice of reduction matrix, $M = V_n^\top$, which minimizes $\|W(I - M^+M)\|_2$ from Theorem S9. Of course, this doesn't mean that it is the reduction matrix that minimizes the alignment error in $\mathbb{R}^n$ (which is another problem in itself), but it provides a reduction matrix that is independent of position x and time t: it solely depends on the structure of the system. The theorem also provides a criterion for exact dimension reduction or in images, perfect alignment of the complete and reduced vector fields, as shown in the following corollary.

Corollary S74. *If all conditions of Assumptions S70 hold, $J_x(x', y') = aI$ for some real constant a, and $n = \operatorname{rank} W$, then the alignment error $\mathcal{E}(x)$ vanishes for all x.*

Proof. Setting $J_x(x', y') = aI$ eliminates the first term of the bound in Theorem S72 for any M:

$$\|MJ_x(x', y')(I - P)x\| = a\|(M - MM^+M)x\| = 0, \tag{S224}$$

since $MM^+M = M$ according to the defining properties of the Moore-Penrose pseudoinverse. Finally, if $n = \operatorname{rank} W$, then $\sigma_{n+1} = 0$, which cancels out the second term of the bound. □

Let s be a vector of N functions $s_i : \mathbb{R} \to \mathbb{R}$. Let W be a $N \times N$ matrix of rank $r < N$ with compact SVD $U_r \Sigma_r V_r^\top$. If $M = V_r^\top$, then by Corollary S74, the dynamics

$$\dot{x} = -d\,x + s(Wx) \tag{S225}$$

can be exactly reduced to the r-dimensional reduced dynamics

$$\dot{X} = -d\,X + V_r^\top s(U_r \Sigma_r X), \qquad X = V_r^\top x\,. \tag{S226}$$

Example S75. The simplest example is the linear dynamics

$$\dot{x} = Wx\,, \tag{S227}$$

where W is not restricted to be the weight matrix in itself and the exact reduction is

$$\dot{X} = V_r^\top U_r \Sigma_r X, \qquad X = V_r^\top x\,. \tag{S228}$$

Example S76. A noteworthy example of dynamics of the form (S225) is the RNN defined by Eqs. (S201). Therefore, when $\operatorname{rank} W = r < N$ and $d_i = d$ for all $i \in \{1, ..., N\}$, the RNN exactly reduces to the r-dimensional dynamics

$$\dot{X} = -dX + V_r^\top \tanh(U_r \Sigma_r X + c), \qquad X = V_r^\top x\,, \tag{S229}$$

where U_r, Σ_r, $V_r^\top$ form the compact SVD of the neural network W.

The RNN used in reservoir computing [121] also involves a dynamics of the form (S201) (with, of course, the important output equation $y = W^{(\text{out})}x$). It can thus be exactly reduced too. Note, however, that the learned matrix W is generally of full rank, but it can have a low effective rank. By shrinking the singular values (with optimal shrinkage [29] for instance) of W, one can get a new RNN and then apply the last result to have a low-dimensional RNN. In other words, one can truncate the neural network W at some rank k —yielding the rank k matrix W_k—in such a way that there is no cost at reducing to k equations the N-dimensional RNN depending on W_k (except the preliminary cost of truncating W).

Example S77. The Wilson-Cowan dynamics in Eq. (S188) with $a = 0$ and $d_j = d$ for all $j \in \{1, ..., N\}$ is essentially equivalent, from a mathematical perspective, to the RNN of the last example. It can thus be exactly reduced to the dynamics

$$\dot{X} = -dX + V_r^\top \mathcal{S}[b(\gamma U_r \Sigma_r X - c)], \qquad X = V_r^\top x\,. \tag{S230}$$

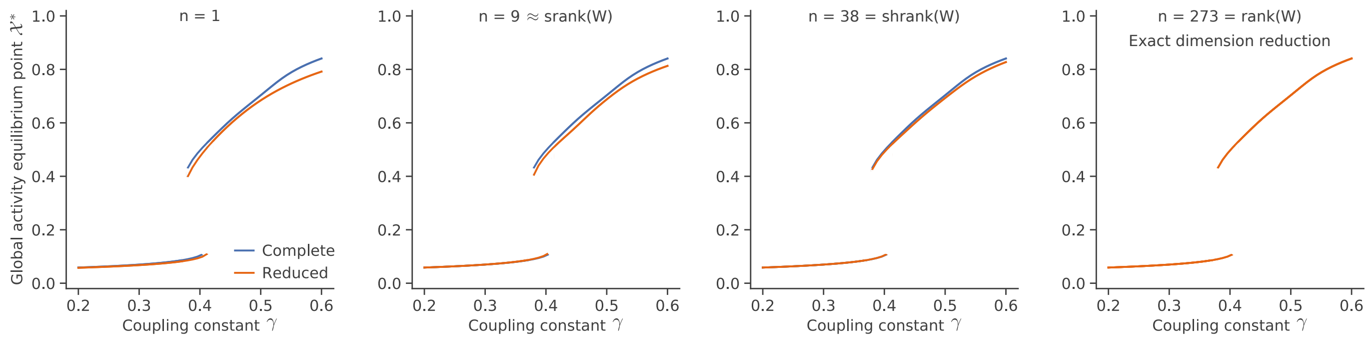

Fig. S8: Comparison between the global observable at equilibrium $\mathcal{X}^*$ of the complete (blue) and reduced (orange) Wilson-Cowan dynamics on the (unsigned) *C. elegans* connectomes ($N = 279$, $\mathrm{rank}(W) = 273$) vs. the global coupling γ for $n \in \{1, 9, 38, 273\}$. Parameters: $d = 1$, $a = 0$, $b = 1$, $c = 3$. For the weight matrix, see the GitHub repository, module get_real_network.py, function get_connectome_weight_matrix (graph_name="celegans"). The effective ranks of this connectome with weight matrix W are $\mathrm{srank}(W) \approx 9$, $\mathrm{thrank}(W) = 27$, $\mathrm{elbow}(W) = 31$, $\mathrm{nrank}(W) \approx 36$, $\mathrm{shrank}(W) = 38$, $\mathrm{energy}(W) = 106$, and $\mathrm{erank}(W) \approx 192$.

In Fig. S8, we illustrate this result for a real connectome by comparing the global observable at equilibrium (see subsection III F) of the complete and reduced dynamics $\dot{X} = -dX + V_n^\top \mathcal{S}[b(\gamma U_n \Sigma_n X - c)]$ with $X = V_n^\top x$ and different values of n.

Example S78. The threshold-linear model in Eq. (S194) with $\mathrm{rank}\, W = r < N$ can also be exactly reduced (despite the discontinuity in the vector field) to the r-dimensional reduced dynamics

$$\dot{X} = -X + V_r^\top \left[U_r \Sigma_r X + b\right]_+ , \qquad X = V_r^\top x\,, \tag{S231}$$

where $X = V_r^\top x$ and U_r, Σ_r, $V_r^\top$ form the compact SVD of the neural network W. To apply Corollary S74, one can simply replace $[\,]_+$ by the softplus function to satisfy the condition 1 of Assumptions S70.

In the case of a linear dynamics, not only the dimension reduction is exact for $n = r$ (Example S75), but the upper bound S214 on the relative alignment error in Theorem S72 takes a very simple form.

Corollary S79. *If the dynamics is a linear system $\dot{x} = Wx$ and Assumptions S70 (2) and (3) are satisfied, the relative alignment error in $\mathbb{R}^n$ at $x \in \mathbb{R}^N$ is*

$$\frac{\mathcal{E}(x)}{\|x\|} \leq \frac{\sigma_{n+1}}{\sqrt{n}}, \tag{S232}$$

where σ_i is the i-th singular value of W.

Proof. It is clearly seen by following the steps of Theorem S72. Indeed, for the linear case, the reduced dynamics is

$$\dot{X} = WM^+X \tag{S233}$$

and the alignment error is

$$\mathcal{E}(x) = \frac{1}{\sqrt{n}}\|M(Wx - WM^+X)\| = \frac{1}{\sqrt{n}}\|MW(I-P)x\|, \tag{S234}$$

where $P = M^+M$. The induced spectral norm and the submultiplicativity imply that

$$\mathcal{E}(x) \leq \frac{1}{\sqrt{n}}\|M\|_2\|W(I-P)\|_2\|x\|. \tag{S235}$$

Assumption S70 (3) then leads to

$$\mathcal{E}(x) \leq \frac{1}{\sqrt{n}}\|W(I - V_nV_n^\top)\|_2\|x\| \tag{S236}$$

and using Theorem S9 gives the desired result. □

For a linear system, the relative alignment error is solely dependent on the $(n+1)$-th singular values and a scaling factor $1/\sqrt{n}$. As a consequence, a rapid decrease of the singular values of W directly induces a rapid decrease of the alignment error.

E. Computation of the upper bound on the alignment error

The bound in Theorem S72 depends on some real point x' which is unknown *a priori*. Yet, according to Eqs. (S216-S218), it is possible to find x' analytically (sometimes exactly) or numerically from

$$[G_x(x') + G_y(x')W]\,(I-P)x = u(x) - u(Px)\,, \tag{S237}$$

where $G_x(x') = J_x(x', y')$, and $G_y(x') = J_y(x', y')$. Below, we give four examples, one for each dynamics used in the paper to produce Fig. 4, from the simplest to the more complex case.

Example S80 (Epidemiological)**.** For the QMF SIS dynamics in Eq. (S192), we can exactly find x'. We have

$$u(x) = -Dx + \gamma(1-x)\circ Wx \tag{S238}$$

$$G_x(x') = -D - \gamma\,\mathrm{diag}(Wx') \tag{S239}$$

$$G_y(x') = \gamma[I - \mathrm{diag}(x')], \tag{S240}$$

where $D = \mathrm{diag}(d_1, ..., d_N)$. By substituting the expressions above in Eq. (S237) and by canceling some terms, we have

$$Wx' \circ \chi + x' \circ W\chi = x \circ Wx - Px \circ WPx. \tag{S241}$$

where $\chi = (I-P)x$. The commutativity of the Hadamard product implies

$$\chi \circ Wx' + W\chi \circ x' = x \circ Wx - Px \circ WPx, \tag{S242}$$

which can be written as a linear equation in x', i.e.,

$$[\mathrm{diag}(\chi)W + \mathrm{diag}(W\chi)]x' = x \circ Wx - Px \circ WPx, \tag{S243}$$

If the matrix $\mathrm{diag}(\chi)W + \mathrm{diag}(W\chi)$ is invertible (which is true in general), then the unique solution to the linear system is

$$x' = [\mathrm{diag}(\chi)W + \mathrm{diag}(W\chi)]^{-1}(x \circ Wx - Px \circ WPx). \tag{S244}$$

In rare cases, if the matrix $\mathrm{diag}(\chi)W + \mathrm{diag}(W\chi)$ is singular, then one can use the least-square optimal solution by using the pseudo-inverse. That being said, using Eq. (S244), one can compute exactly the upper bound on the alignment error for the QMF SIS. In Fig. 4a, we compute the bound for the network of high school contacts from Netzschleuder. For each n and each of the 1000 samples of x with elements between 0 and 1 (the dynamics is bounded between 0 and 1), the diagonal elements in D are sampled from a Gaussian probability density function with mean 1 and standard deviation 0.001 and the coupling constant γ is sampled from a uniform probability density function between 0.01 and 4. In this parameter region, there is a transcritical bifurcation for the global observable defined in subsection III F (see Fig. 4e).

It is sometimes unnecessary to find x' in itself to compute the bound if the Jacobian matrices solely depend on a function of x', as shown in the next example.

Example S81 (RNN)**.** Another way to write the RNN (with no current) is

$$\dot{x}_i = -d_i x_i + \tanh(\gamma \textstyle\sum_{i=1}^N W_{ij}x_j) = -d_i x_i + 2\mathcal{S}(2\gamma \sum_{i=1}^N W_{ij}x_j) - 1. \tag{S245}$$

For the RNN dynamics, we have

$$u(x) = -Dx + \tanh(\gamma Wx) = -Dx + 2\mathcal{S}(2\gamma Wx) - 1 \tag{S246}$$

$$G_x(x') = -D \tag{S247}$$

$$G_y(x') = 4\gamma\,\mathrm{diag}\,[\mathcal{S}(2\gamma Wx')[1 - \mathcal{S}(2\gamma Wx')]] \tag{S248}$$

where $D = \mathrm{diag}(d_1, ..., d_N)$, $\mathcal{S}$ is the sigmoid function. We observe that $G_x(x')$ do not depend over x' and $G_y(x')$ solely depends on the derivative of $\mathcal{S}(2\gamma Wx')$ so we won't have to look for x'. By substituting the expressions above in Eq. (S237) and by canceling some terms, we get

$$\mathcal{S}(2\gamma Wx')[1 - \mathcal{S}(2\gamma Wx')] = \frac{1}{2\gamma}\,\mathrm{diag}[W(I-P)x]^{-1}[\mathcal{S}(2\gamma Wx) - \mathcal{S}(2\gamma WPx)] \tag{S249}$$

which can be directly substituted into $G_y(x')$ to compute the upper bound. In Fig. 4c, we compute the bound for the learned network mouse-control1-model.npz from Ref. [238]. For each n and each of the 1000 uniform samples of x with elements between -1 and 1 (the dynamics is bounded between -1 and 1), the diagonal elements in D are sampled from a Gaussian probability density function with mean 1.6 and standard deviation 0.001 and the coupling constant γ is sampled from a uniform probability density function between 0.16 and 4.8. This parameter region covers convergent and oscillatory dynamics for the RNN. See the script simulations/trajectories_rnn.py on the Github repository low-rank-hypothesis-complex-systems to generate trajectories.

Letting some parameters be small or using numerical optimization, one can get reasonable approximations of the upper bound.

Example S82 (Neuronal)**.** For the Wilson-Cowan dynamics, we have

$$u(x) = -Dx + (1 - ax) \circ \mathcal{S}[b(\gamma W x - c)] \tag{S250}$$
$$G_x(x') = -D - a\,\mathrm{diag}(\mathcal{S}[b(\gamma W x' - c)]) \tag{S251}$$
$$G_y(x') = b\gamma[I - \mathrm{diag}(x')]\,\mathrm{diag}(\mathcal{S}[b(\gamma W x' - c)](1 - \mathcal{S}[b(\gamma W x' - c)])). \tag{S252}$$

From there, various methods can be used to evaluate the upper bound.

1. If $a = 0$, one can get evaluate the upper bound exactly as in the RNN case, with the difference that the sigmoid function depends over the two other parameters b and c. If a is sufficiently close to 0, one can also proceed as in the RNN case, but it will give an approximation of the error bound. In this case, the Jacobian $G_x(x')$, depending on a and appearing in the first term of the error bound, become more and more important *relatively to the second term* as n increases.
2. Instead of trying to solve Eq. (S237) for x', one can naively set x' as x or Px and choose the one that gives the maximum upper bound value on the alignment error. In this case, the approximation of the upper bound is more accurate for larger n since x and Px get closer and x' is a point between them.
3. Numerical optimization, such as a least-squares method, can be used to find x'. This method requires considerably more computational resources, since for each n and each sample in x, one need to solve a high-dimensional optimization problem.

The code and the tests for each case are given in the Python scripts simulations/errors_wilson_cowan.py and tests/test_error_vector_fields_wilson_cowan.py on the Github repository low-rank-hypothesis-complex-systems. In Fig S9, we show the correspondence between the three methods for the *C. elegans* signed network (see graphs/get_connectome_weight_matrix on the GitHub repository of the paper). For each n and each of the uniform samples in x with elements between 0 and 1 (the dynamics is bounded between 0 and 1), the diagonal elements in D are sampled from a Gaussian probability density function with mean 1 and standard deviation 0.001, the parameter a is sampled uniformly between 0.001 and 0.1, the parameter b is sampled uniformly between 0.5 and 2, the parameter c is sampled uniformly between 2 and 4, and the coupling constant γ is sampled from a uniform probability density function between 0.01 and 1. In Fig. 4b, the same parameters as above are used and we apply the second method to get x' since it is faster to compute and it is more precise for large n than the first one.

For some dynamics, it is not trivial to find an approximation like the first method in Example S82 that helps solve Eq. (S237) in x'.

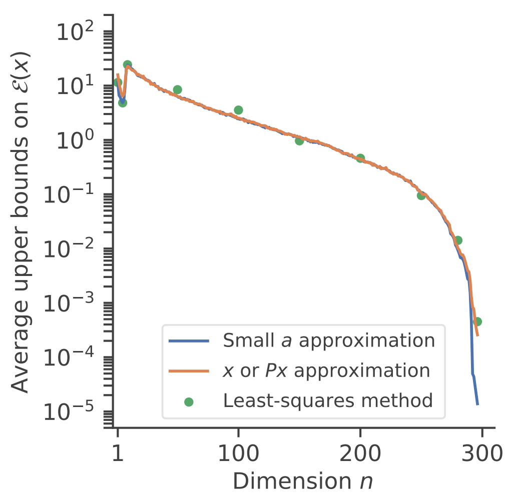

Fig. S9: The upper bounds on the alignment error $\mathcal{E}(x)$ of the neuronal dynamics for different approximation methods of x'. The blue line corresponds to the approximation $a \approx 0$ (1000 samples for each n), the orange line corresponds to the approximation that x' is either x or Px (1000 samples for each n) and the green circles correspond to the approximation of x' using a least-squares method (10 samples for each $n \in \{1, 50, 100, 150, 200, 250, 296\}$).

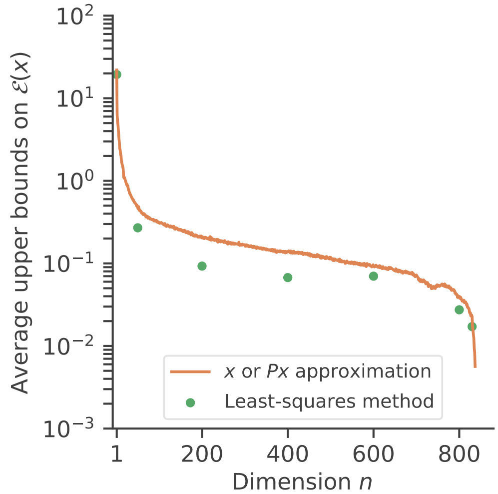

Fig. S10: The upper bounds on the alignment error $\mathcal{E}(x)$ of the microbial dynamics for different approximations of x'. The orange line is related to the approximation that x' is x or Px (1000 samples for each n, 35h of simulations on a personal computer with an Intel i7 processor) and the green circles are related to the approximation of x' using a least-squares method (7 samples for each $n \in \{1, 50, 200, 400, 600, 800, 830\}$).

Example S83 (Microbial)**.** For the microbial population dynamics defined in Eq. (S197), we have

$$u(x) = a - dx + b\,x \circ x - c\,x \circ x \circ x + \gamma\, x \circ Wx \tag{S253}$$
$$G_x(x') = -dI + \operatorname{diag}(2bx' - 3c\,x' \circ x' + \gamma W x') \tag{S254}$$
$$G_y(x') = \gamma \operatorname{diag}(x'). \tag{S255}$$

In matrix form, it is easily shown that the system of equations to solve is

$$A(x' \circ x') + Bx' - C = 0, \tag{S256}$$

where $\chi = (I - P)x$, $D_v = \operatorname{diag}(v_1, ..., v_N)$, and

$$A = -3cD_\chi \tag{S257}$$
$$B = 2bD_\chi + \gamma D_{W\chi} + \gamma D_\chi W \tag{S258}$$
$$C = b\,[x \circ x - Px \circ Px] - c\,[x \circ x \circ x - Px \circ Px \circ Px] + \gamma\,[x \circ (Wx) - (Px) \circ (WPx)]. \tag{S259}$$

In this case, we could not find x' mathematically, since we have to find a root of a system of N coupled quadratic equations, which is a problem in the realm of geometric algebra. Concerning the possibility of making approximations, from our numerical experiments, neither the coupling term $\gamma D_\chi W x'$ nor the quadratic term can be neglected. Moreover, for the parameters $a = 5$, $b = 13$, $c = 10/3$ (or $c = 1$), $d = 30$, $\gamma \in [0.5, 3]$ and the human gut microbiome network [205, 239], the dynamics is not bounded above by 1. Since the alignment error is not a relative error, it can thus take very high values. Even if it's not a problem in itself, to be coherent with the dynamics in the three previous example, we rescale x_i and t in the dynamics to have trajectories approximately bounded between 0 and 1 and to normalize the human gut microbiome network by its largest singular value $\sigma_1 = 171$. To achieve that, we have generated trajectories for the given set of parameters above and we found that the trajectories are (safely) bounded by 30 given, and so we set $x_i \mapsto x_i/30$. Thus, with $t \mapsto 20\,\sigma_1\, t$, we get the differential equations

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = a - d\,x_i + b\,x_i^2 - c\,x_i^3 + \gamma\,x_i \sum_{j=1}^{N} \hat{W}_{ij} x_j, \tag{S260}$$

where the parameters are redefined such that $a \mapsto a/T \approx 5\times10^{-5}$, $d \mapsto d/T \approx 0.01$, $b \mapsto bd/T \approx 0.1$, $c \mapsto cd^2/T \approx 0.9$, $\gamma \mapsto \gamma d/20 \in [0.5, 4.5]$, and $\hat{W} = W/\sigma_1$. In Fig. 4d, we use the second method in Example S82 (x' is x or Px), which is compared to the least-squares method in Fig. S10. Because this is just an approximation of the bound, it is not guaranteed that for a given instance in x and a given n, the value of the bound is above the error, but it is above on average as one can see in Fig. 4d. Also, for each n and each of the uniform samples in x with elements between 0 and 1, $d = 0.01$, the parameter a is sampled uniformly between 0.00001 and 0.0001, the parameter b is sampled uniformly between 0.05 and 2, the parameter c is sampled uniformly between 0.5 and 1.5, and the coupling constant γ is sampled from a uniform probability density function between 0.1 and 5.

F. Global observables

We here describe how to define an observable that describes the activity (state) of dynamics on networks at large scale, allowing the production of a two-dimensional diagram depicting the influence of a structural parameter on the equilibrium states of the (macro-)dynamics.

Numerically, the SVD might give a right singular vector matrix $V = (v_1 \,...\, v_N)$ with many negative entries. For instance, the leading singular vector might contain solely negative elements. Other singular vectors v could be such that $\sum_i v_i < 0$. As a consequence, the dynamics of the observables X_μ (even the leading one, i.e., X_1 related to σ_1) can have equilibrium points below 0, which might be harder to interpret. One way to get more positive values without using any approximation (e.g., nonnegative matrix factorization [154, 240]) is to play with the non-uniqueness of the SVD by multiplying the singular vector matrices by a diagonal matrix $D_\pm$ of +1 and -1. Let

$$D_\pm = \text{diag}\,[\; s(\textstyle\sum_i (v_1)_i) \;,\; ... \;,\; s(\sum_i (v_N)_i) \;]\,, \tag{S261}$$

where s is defined such that

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0. \end{cases} \tag{S262}$$

Since $D_\pm$ is diagonal, it commutes with any diagonal matrices. Moreover, $D_\pm D_\pm = I$. Therefore,

$$W = U\Sigma V^\top = U\Sigma D_\pm D_\pm V^\top = UD_\pm \Sigma D_\pm V^\top := U'\Sigma V'^\top. \tag{S263}$$

To get an approximate reduced dynamics of dimension n, we use the truncated SVD $U_n\Sigma_n V_n^\top$, where $\Sigma_n := \text{diag}(\sigma_1, ..., \sigma_n)$ and the $N \times n$ truncated singular vector matrices are

$$U_n := (u_1' \,...\, u_n') \text{ and } V_n := (v_1' \,...\, v_n'). \tag{S264}$$

After integrating the reduced dynamics with $M = V_n^\top$, we compute the global observable

$$\mathcal{X} = w \cdot X = m \cdot x\,, \tag{S265}$$

where w is a $n \times 1$ vector of constants, $m = w^\top M$ and $\cdot$ is the scalar product. From there, one can define observables on different scales (we will roughly say that we have a global/macroscopic observable if all or almost all vertices contributes to its value through their state). Indeed, depending on w and M, the $1 \times N$ vector $w^\top M$ could have zero elements (say, elements $i, j, ...$), thus canceling the contribution of the activity of some vertices $(x_i, x_j, ...)$ to the observable $\mathcal{X}$. The weight matrices used in the paper (ultimately defining $M = V_n^\top$) and the chosen vector w will lead to global observable, as defined in what follows.

For the epidemiological dynamics, we choose

$$w = (w_1 \; 0 \; ... \; 0)^\top\,, \quad \text{where} \quad w_1 = \frac{1}{\sum_{j=1}(v_1)_j} \tag{S266}$$

with v_1 being the leading right singular vector. This defines, from Eq. S265, the leading right-singular-vector observable

$$\mathcal{X} = \sum_{i=1}^{N} m_i x_i \quad \text{with} \quad m_i = \frac{(v_1)_i}{\sum_{j=1}^{N}(v_1)_j}. \tag{S267}$$

Since the high-school contact network is a nonnegative matrix, it satisfies the Perron-Frobenius theorem and the weight m_i can be interpreted as the hub centrality of vertex i (cf. Fig. S5). The leading right-singular-vector observable hence describes the activity of all the vertices by giving more importance to the ones with high centrality. For the neuronal dynamics, we use the rescaled leading right-singular-vector observable where

$$w = (w_1 \; 0 \; ... \; 0)^\top\,, \quad \text{where} \quad w_1 = \frac{1}{r\sum_{j=1}(v_1)_j} \tag{S268}$$

with $r = 0.15$, to have a bifurcation diagram between 0 and 1 approximately in the range of coupling considered in Fig. 4f.

For the microbial dynamics on the gut microbiome (signed network), a different global observable to show positive, stable equilibrium point branches and to illustrate another way of defining a global observable with our framework. Two criteria are imposed to define the vector m defining the global observable: (1) it does not vary with n and (2) it is as close as possible to the uniform observable where $m_i^{\text{uni}} = 1/(rN)$ for all $i \in \{1, ..., N\}$ and for some rescaling constant r. The first one is imposed strictly while the other is not. To satisfy these conditions, let $n_{\min}$ be the smallest dimension considered (in Fig. 4g, $n_{\min} = 76$) and define the n-dimensional vector

$$w = (w_1 \; ... \; w_{n_{\min}} \; 0 \; ... \; 0)^\top \tag{S269}$$

and the $n_{\min}$-dimensional vector

$$\bar{w} = (w_1 \; ... \; w_{n_{\min}})^\top. \tag{S270}$$

Satisfying condition (2) is equivalent to the problem of finding the coefficient $\bar{w}$ that minimizes $\|V_{n_{\min}}\bar{w} - \mathbb{1}/(rN)\|$ where $\mathbb{1}$ is a N-dimensional vector of ones, which simply gives

$$\bar{w} = \frac{1}{rN} V_{n_{\min}}^\top \mathbb{1}. \tag{S271}$$

where we chose $r = 10$. Condition (1) is thus satisfied and one can compare the equilibrium points of the global observable at different values of $n \geq n_{\min}$ (such as $n \in \{76, 203, 735\}$ in Fig.4).

Note that the above global observables are not chosen in a way that the bifurcation diagram or the trajectories of the complete dynamics are described in the best way possible by the reduced dynamics (in other words, some global observable are better described by the reduced dynamics than others), but rather in a way that they are more intuitive.

G. Numerical efficiency

When integrating the dynamics, the vector field is evaluated many times. For instance, with the integration method DOPRI45, the vector field is evaluated six times at each time step. It is therefore interesting to report a speed comparison for the evaluation of (1) the exact reduced vector field $M \circ h$, (2) the reduced dynamics $M \circ h \circ M^+$, and (3) the reduced dynamics in its tensor form.

TABLE SI: Average time taken to evaluate the exact vector field $M \circ h$, the unsimplified reduced vector field $M \circ h \circ M^+$, and the reduced vector field in closed form with higher-order interactions (closed-tensor form) for different dynamics and different values of n. Parameters: $N = 500$, $x_i \sim \mathcal{U}[0,1)$, $\theta_i, \alpha \sim \mathcal{U}[0, 2\pi)$, $D_{ii} \sim \mathcal{U}[0,1)$, $W_{ij} \sim \mathcal{U}[-1,1)$. The average was computed over 500 time samples of the above parameters. The experiments were done on a basic laptop (Intel i7 MSI GL62 6Qf) and the related Python scripts are gather in the folder tests/test_dynamics of the openly accessible GitHub repository low-rank-hypothesis-complex-systems.

		Average evaluation time [s]		
n	**Reduced vector field**	Lotka-Volterra	QMF SIS	Kuramoto-Sakaguchi
1	Exact $M \circ h$	3.2×10^{-4}	3.3×10^{-3}	9.4×10^{-3}
	Unsimplified $M \circ h \circ M^+$	3.3×10^{-4}	3.1×10^{-3}	9.0×10^{-3}
	Tensor form	1.0×10^{-5}	2.0×10^{-5}	5.6×10^{-5}
10	Exact $M \circ h$	3.6×10^{-4}	3.4×10^{-3}	9.2×10^{-3}
	Unsimplified $M \circ h \circ M^+$	3.0×10^{-4}	2.8×10^{-3}	8.9×10^{-3}
	Tensor form	2.0×10^{-5}	2.9×10^{-5}	9.2×10^{-5}
100	Exact $M \circ h$	5.4×10^{-4}	3.4×10^{-3}	1.0×10^{-2} *
	Unsimplified $M \circ h \circ M^+$	4.6×10^{-4}	3.3×10^{-3}	1.1×10^{-2} *
	Tensor form	8.6×10^{-4}	9.0×10^{-4}	1.2×10^{0} *

*Computed with 10 samples instead of 500.

As shown in Table SI, when we have the argument of each vector field in hand and n is small, there can be significant benefits to use the reduced dynamics in its tensor form (approximately 10-100 times faster than the unsimplified reduced vector field and the complete vector field). The advantage of this reduced dynamics is that the tensors can be computed before the integration of the dynamics. Hence, only quantities depending on n are involved in the integration. For reasonable sizes n, N, and for small enough tensor order, the tensors can be efficiently computed using some tensor calculus or using nested for loops optimized with Numba [see graphs/compute_tensors.py and the speed test in tests/test_graphs/test_compute_tensor.py].

Note, however, that for specific M, W, X, the vector field $M \circ h \circ M^+$ might be faster to evaluate than the closed-tensor form. For large values of n, the tensor form is particularly slow to compute, especially when the order of the tensor is higher (e.g., Kuramoto-Sakaguchi for $n = 100$). The time required to evaluate the unsimplified vector field is more stable according to the size n. We thus extensively used it to compute the alignment error and its upper bound. More exhaustive numerical work should be done in the future to assess the benefits and the limitations of choosing a particular form of the reduced vector field in terms of computation time.

H. Numerical integration of the dynamics

The dynamics on real networks considered in Fig. 4 have very different properties at equilibrium and choosing a correct ordinary differential equation integrator is essential to ensure reliable results. For the epidemiological, neuronal and recurrent neural dynamics, using the algorithm DOPRI45 (see the github repository, dynamics/integrate.py, function integrate_dopri45) to get the equilibrium points of the dynamics worked properly when adjusting the time length and the integration step correctly. For the epidemiological dynamics, the phenomenon of critical slowing down appears, but it can be easily dealt with by increasing the number of time steps near the bifurcation.

The more challenging problem was the integration of the microbial dynamics on the gut microbiome, since the differential equations are stiff: an really small time step for DOPRI45 was needed to capture the very fast transitions in the first steps of the trajectories and the numerical integration was excessively long. Moreover, there are multiple branches of stable equilibrium points close to each other for the global observable (see subsection III F).

We have thus turned to solve_ivp from scipy.integrate with the backward differentiation formula (BDF), an implicit method with variable step length and order. As mentioned in the documentation [https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html] and in Ref. [241], the method is well suited for stiff problems and we have made great computational time gain by using this method since the integrator uses very small steps at the beginning and much larger steps near the equilibrium point. We observed that a relative tolerance of 10^{-8} and an absolute tolerance of 10^{-12} for the complete dynamics and a relative tolerance of 10^{-6} and an absolute tolerance of 10^{-10} for the reduced dynamics were reasonable in terms of integration reliability and computational time for our problem while being in line with the recent benchmarks in Ref. [241]. Moreover, we provided the Jacobian matrices of the complete and reduced dynamics to the integrator as recommended in the documentation of solve_ivp for the BDF method. We have already computed the Jacobian matrix for the complete microbial dynamics to compute the alignment errors in subsection III E, we recall that it is given by

$$Du(x) = G_x(x) + WG_y(x), \tag{S272}$$

where u, G_x and G_y are given in Eqs. (S253-S255) for the microbial dynamics. One can then easily show that the Jacobian matrix of the reduced dynamics with vector field $U = MuM^+$ is

$$DU(X) = M\,Du(M^+X)\,M^+. \tag{S273}$$

In our simulations, we observed that there are many lower (forward) branches of stable equilibrium points near 0 and many other stable equilibrium points (backward) branches at higher values for the global observable. Getting all these branches would be a tremendous challenge and would require sampling an 838-dimensional space of initial conditions, which is far from the goal of the paper. We thus sampled from different initial value uniform distributions to capture some of these branches. We have focused on one forward branch only to illustrate one transition: we observed that sampling the initial condition x_0 from a uniform distribution between 0 and 1 gave only one branch that eventually loses its stability to fall on some other branch at higher activity when increasing the coupling value. To obtain a backward branch, (1) we sampled the initial condition x_0 from a uniform distribution between 0 and z where z is a random integer between 1 and 15, (2) we integrated to get an equilibrium point, (3) we decreased the coupling and used the last equilibrium point as the initial condition for the integration in step (2), and (4) we repeat the steps (2) and (3) until the minimum coupling value (0.1 in Fig. 4g) was reached. We repeated these four steps 100 times (300 for $n = 76$) to generate different initial conditions and different stable branches while ensuring at each iteration that the equilibrium points had reached the tolerance (10^{-7}) and that the equilibrium points were positive. The code to obtain Fig. 4g is on the Github repository, in simulations/bifurcations_microbial.py.

Because of the performances of BDF with the microbial dynamics, we also integrated the other dynamics with the BDF method with a relative tolerance of 10^{-8} and an absolute tolerance of 10^{-12}.

IV. REAL NETWORK DATASET

In this section, we list the real networks used in the paper and we provide two supplementary figures. Every network in the table is from Netzschleuder, except 31 of them, listed below.

- 'celegans_signed': It is obtained by completing (with Dale's principle) the connectome NT+R method prediction of the open-source database EleganSign [242] [see graphs/get_real_networks.py, function get_connectome_weight_matrix in the GitHub repository].
- 'drosophila': It is taken from Ref. [243].
- 'cintestinalis' The Ciona intestinalis connectome is from Ref. [244] and is available on our Github repository in graphs/graph_data/connectome/ciona_intestinalis_lavaire_elife-16962-fig16-data1-v1_modified.xlsx.
- 'pdumerilii_neuronal': The neuronal Platynereis dumerilii connectome is from Ref. [245] and it is an updated version shared personally by the author G. Jékely to V. Thibeault. The connectome is available on our Github repository in graphs/graph_data/connectome/pdumerilii_neuronal.xml.
- 'pdumerilii_desmosomal': The desmosomal Platynereis dumerilii connectome is from Ref. [246] and it is an updated version shared personally by the author G. Jékely to V. Thibeault. The connectome is available on our Github repository in graphs/graph_data/connectome/pdumerilii_desmosomal.xml.
- 'mouse_meso': The mesoscopic mouse connectome is given in Ref. [247] and available on our Github repository in graphs/graph_data/connectome/mouse_connectome-Oh_Nature_2014.csv.
- 'zebrafish_meso': The zebrafish mesoscopic connectome is adapted from Ref. [148] and the treatment is available on the paper's GitHub repository low-rank-hypothesis-complex-systems.
- 'mouse_voxel': The mouse connectome at the level of voxels is available in Mendeley data mouse_connectome_voxelwise [248].
- 'mouse_control_rnn', 'mouse_rnn', 'zebrafish_rnn': recurrent neural networks from Hadjiabadi et al. [238].
- 'fully_connected_layer_cnn_XXXXX' with XXXXX in $\{00100, 00200, ..., 01000\}$: fully connected layers from the convolutional neural networks in the repository NWS[127].
- 'gut': The human gut microbiome is from Ref. [239] and was constructed as in the supplementary material of Ref. [205] [see graphs/get_real_networks.py, function get_microbiome_weight_matrix in the GitHub repository].
- 'AT_2008', 'CY_2015', 'EE_2010', 'PT_2009', 'SI_2016': Economic networks from Ref. [249].
- 'financial_institution07-Apr-1999', 'non_financial_institution04-Jan-2001', 'households_04-Sep-1998', 'households_09-Jan-2002': Economic networks from Ref. [250] on Dryad.

The code to extract each network made available on Github is in graphs/get_real_networks. Other information about the real networks in the dataset is available on the Github repository low-rank-hypothesis-complex-systems. In particular, see real_networks_and_their_effective_ranks.pdf on in graphs/graph_data for the source of each network or equivalently, Supplementary Table 1 (supplementary_table_1_real_networks.pdf). Note that, in a preliminary treatment before getting the effective ranks, many Netzschleuder's networks have been removed from a larger dataset of 1145 networks to avoid over-representation of particular types of networks (specifically, 'board_directors_net1m...', 'edit_wikibooks...', 'ego_social_gplus...').

In subsection II C, asymptotic results about the effective ranks of graph models have been presented for different singular value decays, showing all sorts of behavior, ranging from constant $O(1)$, to sub-linear $O(N^{1-\epsilon})$ with $0 < \epsilon < 1$, to linear $O(N)$ growth as $N \to \infty$. Although we do not expect one graph model to describe every network in the dataset (which would allow doing asymptotic analysis), we can still wonder how the effective ranks are distributed according to the size N of the networks. In Fig. S11, we present such distributions and perform nonlinear regressions, which suggest sub-linear increases of the effective ranks as N increases. As mentioned in subsection II E, it would be pertinent to explore the behavior of the effective ranks in growing graphs and real growing networks to verify the presence of sub-linear growth.

Moreover, sparse matrices have been observed for many real and synthetic networks and in subsection II C, it was shown that sparse matrix models lead to a low stable rank. Yet, Fig. S12 illustrates that the effective ranks are rather anti-correlated with the density of the weight matrices of real networks, thus suggesting that it is really the rapid decrease of the singular values that lead to our observations on the effective ranks in Fig. 1.

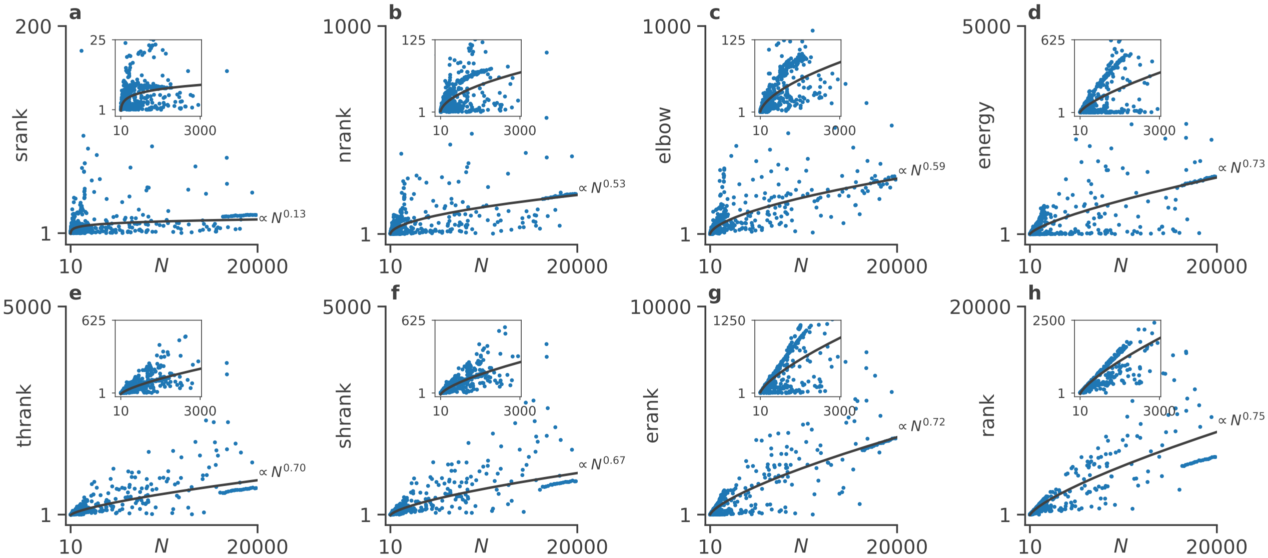

Fig. S11: Different effective ranks vs. the number of vertices N for 679 real networks (see SI IV). The solid black lines are L1 nonlinear regressions with the function $aN^b + c$ and a, b, c as the optimization variables. The insets show zoomed version of the data for the smaller values of effective ranks and N where the nonlinearity is better seen especially in **a** to **d**. The optimization was performed with the method BFGS of scipy.optimize.minimize with the bounds (0, 10), (0, 1), (-100, 30) for a, b, c respectively and the initial guesses $a_{\text{guess}} = 1$, $b_{\text{guess}} = 0.5$, and $c_{\text{guess}} = -1$ (see plot_fig_SI_effective_rank_vs_size.py on the Github repository). The L1 norm was chosen for its better robustness to outliers, but the conclusions hold when using the L2 norm instead. From srank to rank, the optimization parameters $[a\ b\ c]$ are approximately [6.09, 0.13, -7.45], [1.04 0.53 -1.00], [0.76 0.59 -1.00], [1.02 0.73 -1.01], [0.78 0.70 -11.97], [1.27 0.67 -17.44], [2.92 0.72 -0.83], [4.80 0.75 -51.34] and the normalized mean absolute errors $\sum_{i=1}^{679} |y_i - \hat{y}_i|/[679\langle y\rangle]$ are 0.76, 0.69, 0.54, 0.72, 0.51, 0.49, 0.45, 0.41.

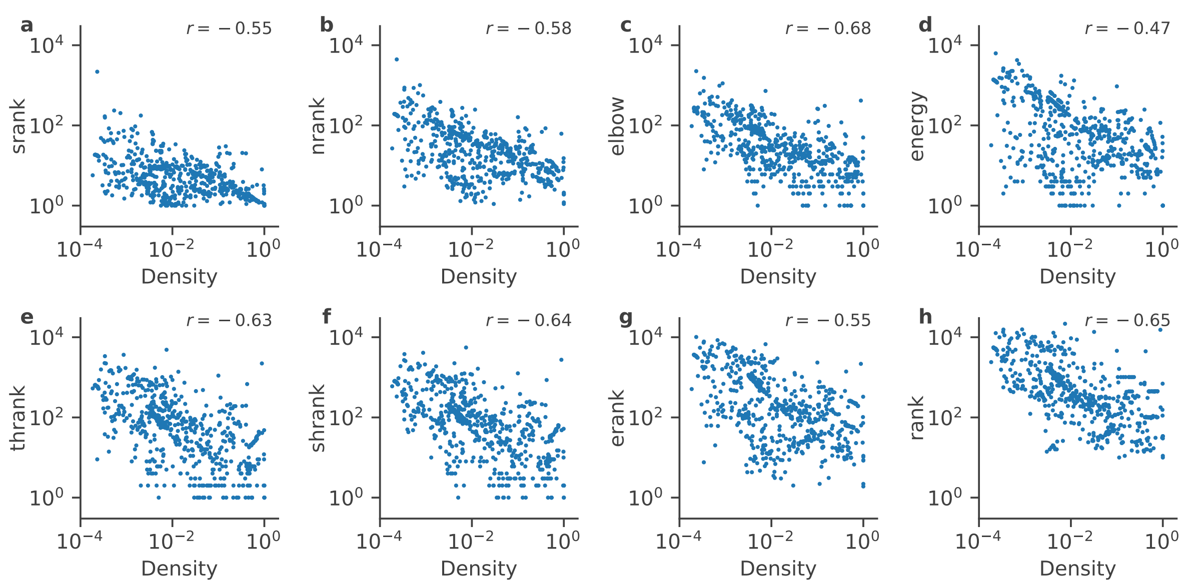

Fig. S12: Different effective ranks vs. the density for 679 real networks (see SI IV). The (matrix) density is the number of nonzero elements in the weight matrices of the networks divided by the total number of elements N^2. The parameter r denotes Pearson's correlation coefficient between the log of the effective ranks and the log of the density.

[1] E. Schmidt, "Zur Theorie der linearen und nichtlinearen lntegralgleichungen." Math. Ann. **63**, 433 (1907).
[2] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," Psychometrika **1**, 211 (1936).
[3] G. W. Stewart, "On the early history of singular value decomposition," SIAM Rev. **35**, 551 (1993).
[4] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, 2019).
[5] J. J. Gerbrands, "On the relationships between SVD, KLT and PCA," Pattern Recognit. **14**, 375 (1981).
[6] H. Hotelling, "Analysis of a complex of statistical variables into principal components," J. Educ. Psych. **24**, 417 (1933).
[7] H. Hotelling, "Analysis of a complex of statistical variables into principal components," J. Educ. Psych. **24**, 498 (1933).
[8] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," Chemom. Intell. Lab. Syst. **2**, 37 (1987).
[9] L. Ferré, "Selection of components in principal component analysis: A comparison of methods," Comput. Stat. Data Anal. **19**, 669 (1995).
[10] H. Abdi and L. J. Williams, "Principal component analysis," WIREs Comput. Stat. **2**, 433 (2010).
[11] I. M. Johnstone and D. Paul, "PCA in High Dimensions: An Orientation," Proc. IEEE **106**, 1277 (2018).
[12] R. D. Cook, "A slice of multivariate dimension reduction," J. Multivar. Anal. **188**, 104812 (2022).
[13] K. Karhunen, *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, Ph.D. thesis, University of Helsinki (1947).
[14] M. Loève, *Probability theory: foundations, random sequences* (Springer, 1955).
[15] R. Everson and L. Sirovich, "Karhunen-Lòeve procedure for gappy data," J. Opt. Soc. Am. A **12**, 1657 (1995).
[16] G. Kerschen, J.-C. Golinval, A. F. Vakakis, and L. A. Bergman, "The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview," Nonlinear Dyn. **41**, 147 (2005).
[17] S. Volkwein, "Proper Orthogonal Decomposition: Theory and Reduced-Order Modelling," (2013).
[18] J. N. Kutz, S. L. Brunton, and B. W. Brunton, *Dynamic Mode Decomposition* (SIAM, 2016).
[19] E. Lorenz, *Empirical Orthogonal Functions and Statistical Weather Prediction*, Tech. Rep. (Massachusetts Institute of Technolog, 1956).
[20] A. H. Monahan, J. C. Fyfe, M. H. P. Ambaum, D. B. Stephenson, and G. R. North, "Empirical orthogonal functions: The medium is the message," J. Clim. **22**, 6501 (2009).
[21] H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," Biol. Cybern. **59**, 291 (1988).
[22] H. Bourlard and S. H. Kabil, "Autoencoders reloaded," Biol. Cybern. **116**, 389 (2022).
[23] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. (Springer, New York, 2010).
[24] F. Benaych-Georges and R. R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," J. Multivar. Anal. **111**, 120 (2012).
[25] T. Tao, *Topics in Random Matrix Theory*, Vol. 132 (American Mathematical Society, 2012).
[26] T. Tao and V. Vu, "Random covariance matrices: Universality of local statistics of eigenvalues," Ann. Probab. **40**, 1285 (2012).
[27] M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is $4/\sqrt{3}$," IEEE Trans. Inf. Theory **60**, 5040 (2014).
[28] A. Bloemendal and B. Virág, "Limits of spiked random matrices ii," Ann. Probab. **44**, 2726 (2016).
[29] M. Gavish and D. L. Donoho, "Optimal Shrinkage of Singular Values," IEEE Trans. Inf. Theory **63**, 2137 (2017).
[30] B. Beckermann and A. Townsend, "On the singular values of matrices with displacement structure," SIAM J. Matrix Anal. Appl. **38**, 1227 (2017).
[31] D. Donoho, M. Gavish, and I. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model1," Ann. Statis. **46**, 1742 (2018).
[32] P. J. Forrester, *Log-Gases and Random Matrices* (Princeton University Press, 2010).
[33] R. A. Horn and C. R. Johnson, *Topics in matrix analysis* (Cambridge University Press, 1991).
[34] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, 2013).
[35] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," Proc. Natl. Acad. Sci. U.S.A. **37**, 760 (1951).
[36] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities : Theory of Majorization and its Application*, 2nd ed. (Springer, 2011).
[37] H. Wittmeyer, "Einfluiß der Änderung einer Matrix auf die Lösung des zugehörigen Gleichungssystems, sowie auf die charakteristischen Zahlen und die Eigenvektoren." Z. Angew. Math. Mech. **16**, 287 (1936).
[38] L. Mirsky, "Symmetric gauge functions and unitarily invariant norms," Q. J. Math. **11**, 50 (1960).
[39] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed. (Springer, New York, 2003).
[40] A. C. Antoulas, *Approximation of Large-Scale Dynamical System* (SIAM, 2005).
[41] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. (John Hopkins University Press, 2013).
[42] I. Markovsky, *Low-Rank Approximations: Algorithms, Implementation, Applications*, 2nd ed. (Springer, 2019).
[43] N. Harvey, "Low-rank approximation of matrices," (2011), Lecture 15, Section 1, University of British Columbia.
[44] R. Penrose, "Generalized inverse matrices," Math. Proc. Camb. Philos. Soc. **51**, 406 (1955).
[45] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science* (Cambridge University Press, New York, 2018).

[46] M. Rudelson and R. Vershynin, "Sampling from large matrices: An approach through geometric functional analysis," J. ACM **54**, 1 (2007).
[47] M. B. Cohen, J. Nelson, and D. P. Woodruff, "Optimal approximate matrix product in terms of stable rank," in *43rd Int. Colloq. Autom. Lang. Program. (ICALP 2016)*, Vol. 55 (2016) p. 11.
[48] B. Désy, P. Desrosiers, and A. Allard, "Dimension matters when modeling network communities in hyperbolic spaces," arXiv:2209.09201 (2023).
[49] A. Kyrillidis, M. Vlachos, and A. Zouzias, "Approximate matrix multiplication with application to linear embeddings," IEEE Int. Symp. Inf. Theory , 2182 (2014).
[50] I. Gutman, "The energy of a graph: Old and new results," Algebr. Comb. Appl. , 196 (2001).
[51] V. Nikiforov, "The energy of graphs and matrices," J. Math. Anal. Appl. **326**, 1472 (2007); B. Nica, *A Brief Introduction to Spectral Graph Theory* (European Mathematical Society, Zurich, 2018).
[52] A. A. Shabalin and A. B. Nobel, "Reconstruction of a low-rank matrix in the presence of Gaussian noise," J. Multivar. Anal. **118**, 67 (2013).
[53] O. Roy and M. Vetterli, "The effective rank: A measure of effective dimensionality," in *Eur. Signal Process. Conf.* (2007) p. 606.
[54] R. Cangelosi and A. Goriely, "Component retention in principal component analysis with application to cDNA microarray data," Biol. Direct **2**, 1 (2007).
[55] O. Alter, P. O. Brown, and D. Botstein, "Singular value decomposition for genome-Wide expression data processing and modeling," Proc. Natl. Acad. Sci. U.S.A. **97**, 10101 (2000).
[56] L. L. Campbell, "Minimum coefficient rate for stationary random processes," Inf. Control **3**, 360 (1960).
[57] P. O. Perry, *Cross-Validation for Unsupervised Learning*, Ph.D. thesis, Stanford University (2009).
[58] W. Leeb, "Optimal singular value shrinkage for operator norm loss: Extending to non-square matrices," Stat. Probab. Lett. **186**, 109472 (2022).
[59] M. W. Mahoney, "Randomized algorithms for matrices and data," Found. Trends Mach. Learn. **3**, 123 (2011).
[60] P. Gao and S. Ganguli, "On simplicity and complexity in the brave new world of large-scale neuroscience," Curr. Opin. Neurobiol. **32**, 148 (2015).
[61] P. D. Killworth and H. R. Bernard, "Informant accuracy in social network data," Hum. Organ. **35**, 269 (1976).
[62] T. P. Peixoto, "Reconstructing networks with unknown and heterogeneous erors," Phys. Rev. X **8**, 041011 (2018).
[63] M. E. J. Newman, "Network structure from rich but noisy data," Nat. Phys. **14**, 542 (2018).
[64] J.-G. Young, G. T. Cantwell, and M. E. Newman, "Bayesian inference of network structure from unreliable data," J. Complex Netw. **8**, 1 (2020).
[65] J.-G. Young, F. S. Valdovinos, and M. E. J. Newman, "Reconstruction of plant–pollinator networks from observational data," Nat. Commun. **12**, 1 (2021).
[66] Z. Furedi and J. Komlos, "The eigenvalues of random symmetric matrices," Combinatorica **1**, 233 (1981).
[67] P. Bonacich, "Power and Centrality: A Family of Measures," Am. J. Sociol. **92**, 1170 (1987).
[68] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, "Chaos in random neural networks," Phys. Rev. Lett. **61**, 259 (1988); F. Chung, *Spectral Graph Theory* (CBMS, Rhode Island, 1994).
[69] F. Chung, L. Lu, and V. Vu, "Spectra of random graphs with given expected degrees," Proc. Natl. Acad. Sci. U.S.A. **100**, 6313 (2003).
[70] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, and A. N. Samukhin, "Spectra of complex networks," Phys. Rev. E **68**, 046109 (2003).
[71] P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, 2011).
[72] F. Chung and M. Radcliffe, "On the spectra of general random graphs," Electron. J. Comb. **18**, P215 (2011).
[73] R. R. Nadakuditi and M. E. J. Newman, "Graph Spectra and the Detectability of Community Structure in Networks," Phys. Rev. Lett. **108**, 188701 (2012).
[74] T. P. Peixoto, "Eigenvalue Spectra of Modular Networks," Phys. Rev. Lett. **111**, 098701 (2013).
[75] C. Castellano and R. Pastor-Satorras, "Topological determinants of complex networks spectral properties: structural and dynamical effects," Phys. Rev. X **7**, 041024 (2017).
[76] M. E. J. Newman, X. Zhang, and R. R. Nadakuditi, "Spectra of random networks with arbitrary degrees," Phys. Rev. E **99**, 042309 (2019).
[77] A. Athreya, J. Cape, and M. Tang, "Eigenvalues of stochastic blockmodel graphs and random graphs with low-rank edge probability matrices," Sankhya A **84**, 36 (2022).
[78] E. Estrada and P. Knight, *A first course on network science* (Oxford University Press, 2015); A.-L. Barabási, *Network science* (Cambridge University Press, 2016); V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, 2017); M. E. J. Newman, *Networks* (Oxford University Press, 2018).
[79] D. M. Cvetkovic, M. Doob, and H. Sachs, "Spectra of graphs. Theory and application," (1980).
[80] R. Solomonoff and A. Rapoport, "Connectivity of random nets," Bull. Math. Biophys. **13**, 59 (1951).
[81] E. N. Gilbert, "Random graphs," Ann. Math. Stat. **30**, 1141 (1959).
[82] P. Erdős and A. Rényi, "On the evolution of random graphs," Publ. Math. Inst. Hung. Acad. Sci **5**, 17 (1960).
[83] M. E. J. Newman, "The structure and function of complex networks," SIAM Rev. **45**, 167 (2003).
[84] A. Guionnet, "Bernoulli Random Matrices," arXiv:2112.05506 (2021).
[85] A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, "Optimality and sub-optimality of PCA I: Spiked random matrix models," Ann. Statis. **46**, 2416 (2018).

[86] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," Soc. Netw. **5**, 109 (1983).
[87] J.-G. Young, G. St-Onge, P. Desrosiers, and L. J. Dubé, "Universality of the stochastic block model," Phys. Rev. E **98**, 032309 (2018).
[88] F. Chung and L. Lu, "Connected Components in Random Graphs with Given Expected Degree Sequences," Ann. Comb. **6**, 125 (2002).
[89] F. Chung and L. Lu, "The average distances in random graphs with given expected degrees," Proc. Natl. Acad. Sci. U.S.A. **99**, 15879 (2002).
[90] W. Wang, M. Tang, E. H. Stanley, and L. A. Braunstein, "Unification of theoretical approaches for epidemic spreading on complex networks," Rep. Prog. Phys. **80**, 036603 (2017).
[91] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Critical phenomena in complex networks," Rev. Mod. Phys. **80**, 1275 (2008).
[92] E. Valdano and A. Arenas, "Exact rank reduction of network models," Phys. Rev. X **9**, 031050 (2019).
[93] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná, "Hyperbolic geometry of complex networks," Phys. Rev. E **82**, 036106 (2010).
[94] A. Allard, M. Á. Serrano, and M. Boguñá, "Geometric description of clustering in directed networks," arXiv:2302.09055 (2023).
[95] J. Gower, "Properties of Euclidean and non-Euclidean distance matrices," Linear Algebra Appl. **67**, 81 (1985).
[96] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science **289**, 509 (1999).
[97] D. de Solla Price, "A general theory of bibliometric and other cumulative advantage processes," J. Am. Soc. Inf. Sci. **27**, 292 (1976).
[98] C. Aicher, A. Z. Jacobs, and A. Clauset, "Learning latent block structure in weighted networks," J. Complex Netw. **3**, 221 (2015).
[99] T. L. J. Ng and T. B. Murphy, "Weighted stochastic block model," Statistical Methods & Applications **30**, 1365 (2021).
[100] U. Brandes, J. Lerner, U. Nagel, and B. Nick, "Structural trends in network ensembles," in *Complex Networks: Results of the 2009 International Workshop on Complex Networks (CompleNet 2009)* (2009) p. 83.
[101] M. Porfiri, D. J. Stilwell, and E. M. Bollt, "Synchronization in random weighted directed networks," IEEE Transactions on Circuits and Systems I **55**, 3170 (2008).
[102] K. Rajan and L. F. Abbott, "Eigenvalue spectra of random matrices for neural networks," Phys. Rev. Lett. **97**, 188104 (2006).
[103] J. Kadmon and H. Sompolinsky, "Transition to chaos in random neuronal networks," Phys. Rev. X **5**, 041030 (2015).
[104] T. Tao and V. Vu, "Random matrices: the circular law," Communications in Contemporary Mathematics **10**, 261 (2008).
[105] F. Götze and A. Tikhomirov, "The circular law for random matrices," Ann. Probab. **38**, 1444 (2010).
[106] K. P. Costello and V. Vu, "On the rank of random sparse matrices," Combinatorics, Probability and Computing **19**, 321 (2010).
[107] P. M. Wood, "Universality and the circular law for sparse random matrices," Ann. Appl. Probab. **22**, 1266 (2012).
[108] N. Cook, "The circular law for random regular digraphs with random edge weights," Random Matrices: Theory and Applications **6**, 1750012 (2017).
[109] A. Allard, M. Á. Serrano, G. García-Pérez, and M. Boguñá, "The geometric nature of weights in real complex networks," Nat. Commun. **8**, 14103 (2017).
[110] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks," Phys. Rev. E **83**, 016107 (2011).
[111] T. P. Peixoto, "Nonparametric weighted stochastic block models," Phys. Rev. E **97**, 012306 (2018).
[112] A. Athreya, D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, Y. Qin, and D. L. Sussman, "Statistical inference on random dot product graphs: A survey," J. Mach. Learn. Res. **18**, 1 (2018).
[113] D. Garlaschelli and M. I. Loffredo, "Generalized bose-fermi statistics and structural correlations in weighted networks," Phys. Rev. Lett. **102**, 038701 (2009).
[114] D. Garlaschelli, "The weighted random graph model," New J. Phys. **11**, 073005 (2009).
[115] M. A. Serrano and M. Boguñá, "Weighted configuration model," in *AIP conference proceedings*, Vol. 776 (American Institute of Physics, 2005) p. 101.
[116] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature **393**, 440 (1998).
[117] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," Phys. Rev. Lett. **35**, 1792 (1975).
[118] P. Desrosiers and P. Forrester, "Asymptotic correlations for Gaussian and Wishart matrices with external source," Int. Math. Res. Not. **2006**, 27395 (2006).
[119] A. Bloemendal and B. Virág, "Limits of spiked random matrices i," Probab. Theory Relat. Fields **156**, 795 (2013).
[120] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities." Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[121] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," Comput. Sci. Rev. **3**, 127 (2009).
[122] D. Sussillo and L. F. Abbott, "Generating coherent patterns of activity from chaotic neural networks," Neuron **63**, 544 (2009).
[123] F. Mastrogiuseppe and S. Ostojic, "Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks," Neuron **99**, 609 (2018).
[124] F. Schuessler, A. Dubreuil, F. Mastrogiuseppe, S. Ostojic, and O. Barak, "Dynamics of random recurrent networks with

correlated low-rank structure," Phys. Rev. Research **2**, 013111 (2020).
[125] M. Beiran, A. Dubreuil, A. Valente, F. Mastrogiuseppe, and S. Ostojic, "Shaping dynamics with multiple populations in low-rank recurrent networks," Neural Comput. **33**, 1572 (2021).
[126] F. Schuessler, F. Mastrogiuseppe, A. Dubreuil, S. Ostojic, and O. Barak, "The interplay between randomness and structure during learning in RNNs," in *Adv. Neural Inf. Process. Syst. 34* (2020) p. 1.
[127] G. Eilertsen, D. Jönsson, T. Ropinski, J. Unger, and A. Ynnerman, "Classifying the classifier: dissecting the weight space of neural networks," Proceedings of the European Conference on Artificial Intelligence (ECAI 2020) **325**, 1119 (2020).
[128] C. H. Martin and M. W. Mahoney, "Implicit self-regularization in deep neural networks: Evidence from random matrix theory and implications for learning," J. Mach. Learn. Res. **22**, 1 (2021).
[129] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," Ann. Probab. **33**, 1643 (2005).
[130] E. T. Jaynes, "Information Theory and Statistical Mechanics," The Phys. Rev. **106**, 620 (1957).
[131] J. Park and M. E. J. Newman, "Statistical mechanics of networks," Phys. Rev. E **70**, 066117 (2004).
[132] G. Bianconi, "Entropy of network ensembles," Phys. Rev. E **79**, 036114 (2009).
[133] T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics* (Springer, 2017).
[134] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, "The statistical physics of real-world networks," Nat. Rev. Phys. **1**, 58 (2019).
[135] C. Carathéodory, "The beginning of research in the calculus of variations," Osiris **3**, 224 (1937).
[136] G. Giorgi and T. H. Kjeldsen, eds., *Traces and emergence of nonlinear programming* (Birkhauser, New York, 2014).
[137] C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the first order*, 3rd ed. (Chelsea Publishing Company, 1989).
[138] B. H. Pourciau, "Modern Multiplier Rules," Am. Math. Mon. **87**, 433 (1980).
[139] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, 4th ed. (Wiley, New Jersey, 2013).
[140] DLMF, "*NIST Digital Library of Mathematical Functions*," `https://dlmf.nist.gov/`, Release 1.1.9 of 2023-03-15, f. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
[141] J. Gao, Y. Cao, and J.-M. Lee, "Principal component analysis of $1/f\alpha$ noise," Phys. Lett. A **314**, 392–400 (2003).
[142] M. Sánchez-Islas, J. C. Toledo-Roy, and A. Frank, "Criticality in a multisignal system using principal component analysis," Phys. Rev. E **103**, 042111 (2021).
[143] C. Stringer, M. Pachitariu, N. Steinmetz, M. Carandini, and K. D. Harris, "High-dimensional geometry of population responses in visual cortex," Nature **571**, 361 (2019).
[144] C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, and K. D. Harris, "Spontaneous behaviors drive multidimensional, brainwide activity," Science **364**, eaav7893 (2019).
[145] N. C. L. Kong, E. Margalit, J. L. Gardner, and A. M. Norcia, "Increasing neural network robustness improves match to macaque v1 eigenspectrum, spatial frequency preference and predictivity," PLoS Comput. Biol. **18**, e1009739 (2022).
[146] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," in *SODA '98: Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms* (1998) p. 668.
[147] M. E. J. Newman, *Networks* (Oxford University Press, 2018).
[148] M. Kunst, E. Laurell, N. Mokayes, A. Kramer, F. Kubo, A. M. Fernandes, D. Förster, M. Dal Maschio, and H. Baier, "A Cellular-Resolution Atlas of the Larval Zebrafish Brain," Neuron **103**, 21 (2019).
[149] M. Mitchell, *Complexity: A Guided Tour* (Oxford University Press, 2009).
[150] D. Witvliet, B. Mulcahy, J. K. Mitchell, Y. Meirovitch, D. R. Berger, Y. Wu, Y. Liu, W. X. Koh, R. Parvathala, D. Holmyard, R. L. Schalek, N. Shavit, A. D. Chisholm, J. W. Lichtman, A. D. T. Samuel, and M. Zhen, "Connectomes across development reveal principles of brain maturation," Nature **596**, 257 (2021).
[151] K. Fujimoto, "What are singular values of nonlinear operators?" in *43rd IEEE Conf. Decis. Control* (2004) p. 1623.
[152] O. Koch and C. Lubich, "Dynamical low-rank approximation," SIAM J. Matrix Anal. Appl. **29**, 434 (2007).
[153] P. Holme and J. Saramäki, "Temporal networks," Phys. Rep. **519**, 97 (2012).
[154] V. Thibeault, G. St-Onge, L. J. Dubé, and P. Desrosiers, "Threefold way to the dimension reduction of dynamics on networks: An application to synchronization," Phys. Rev. Research **2**, 043215 (2020).
[155] V. Thibeault, *Réduire la dimension des systèmes complexes : un regard sur l'émergence de la synchronisation*, Master's thesis, Université Laval (2020).
[156] C. Kuehn and C. Bick, "A universal route to explosive phenomena," Sci. Adv. **7**, 1 (2021).
[157] A. N. Montanari, C. Duan, L. A. Aguirre, and A. E. Motter, "Functional observability and target state estimation in large-scale networks," Proc. Natl. Acad. Sci. U.S.A. **119**, e2113750119 (2022).
[158] X. Wang and I. H. Sloan, "Why are high-dimensional finance problems often of low effective dimension?" SIAM J. Comput **27**, 159 (2005).
[159] P. Español, "Statistical Mechanics of Coarse-graining," in *Nov. Methods Soft Matter Simulations*, Vol. 140, edited by M. Karttunen, I. Vattulainen, and A. Lukkarinen (Springer, Berlin, 2003) p. 69.
[160] P. Castiglione, M. Falcioni, A. Lesne, and A. Vulpiani, *Chaos and Coarse Graining in Statistical Mechanics* (Cambridge University Press, 2008).
[161] Y. S. Cho, T. Nishikawa, and A. E. Motter, "Stable chimeras and independently synchronizable clusters," Phys. Rev. Lett. **119**, 084101 (2017).
[162] L. D. Smith and G. A. Gottwald, "Model reduction for the collective dynamics of globally coupled oscillators: From finite

networks to the thermodynamic limit," Chaos **30**, 093107 (2020).
[163] J. Wei and J. C. W. Kuo, "A lumping analysis in monomolecular reaction systems: Analysis of the exactly lumpable system," Ind. Eng. Chem. Fundamen. **8**, 114 (1969).
[164] J. Tóth, G. Li, H. Rabitz, and A. S. Tomlin, "The effect of lumping and expanding on kinetic differential equations," SIAM J. Appl. Math. **57**, 1531 (1997).
[165] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of epidemics on networks: From exact to approximate models* (Springer, Cham, 2017).
[166] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, "Parameter Space Compression Underlies Emergent Theories and Predictive Models," Science **342**, 604 (2013).
[167] T. Hoefler, D. Alistarh, T. Ben-Nun, N. Dryden, and A. Peste, "Sparsity in Deep Learning : Pruning and growth for efficient inference and training in neural networks," J. Mach. Learn. Res. **23**, 1 (2021).
[168] F. Forni and R. Sepulchre, "Differential Dissipativity Theory for Dominance Analysis," IEEE Trans. Autom. Control **64**, 2340 (2019).
[169] M. Faccin, M. T. Schaub, and J. C. Delvenne, "State Aggregations in Markov Chains and Block Models of Networks," Phys. Rev. Lett. **127**, 078301 (2021).
[170] J. C. W. Kuo and J. Wei, "A lumping analysis in monomolecular reaction systems: Analysis of the approximately lumpable system," Ind. Eng. Chem. Fundamen. **8**, 124 (1969).
[171] L. Perko, *Differential Equations and Dynamical Systems* (Springer, 2001).
[172] R. Penrose, "On best approximate solutions of linear matrix equations," Math. Proc. Camb. Philos. Soc. **52**, 17 (1956).
[173] J. J. Sylvester, "Sur l'équation en matrices $px = xq$," C. R. Acad. Sci. Paris **99**, 67 (1884).
[174] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 1998).
[175] G. Li and H. Rabitz, "A general analysis of exact lumping in chemical kinetics," Chem. Eng. Sci. **44**, 1413 (1989).
[176] G. Li and H. Rabitz, "A general analysis of approximate lumping in chemical kinetics," Chem. Eng. Sci. **45**, 977 (1990).
[177] C. W. Rowley, T. Colonius, and R. M. Murray, "Model reduction for compressible flows using POD and Galerkin projection," Physica D **189**, 115 (2004).
[178] K. Carlberg, M. Barone, and H. Antil, "Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction," J. Comput. Phys. **330**, 693 (2017).
[179] Editors, "Georgii Ivanovich Petrov (on his 100th birthday)," Fluid Dyn. **47**, 289 (2012).
[180] M. J. Gander and G. Wanner, "From Euler, Ritz, and Galerkin to modern computing," SIAM Rev. **54**, 627 (2012).
[181] Y. Saad, *Iterative methods for sparse linear systems* (SIAM, Philadelphia, 2003).
[182] H. R. Wilson and J. D. Cowan, "Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons," Biophys. J. **12**, 1 (1972).
[183] A. Destexhe and T. J. Sejnowski, "The wilson–cowan model, 36 years later," Biol. Cybern. **101**, 1 (2009).
[184] C. C. Chow and Y. Karimipanah, "Before and beyond the wilson–cowan equations," J. Neurophysiol. **123**, 1645 (2020).
[185] V. Painchaud, N. Doyon, and P. Desrosiers, "Beyond Wilson-Cowan dynamics: oscillations and chaos without inhibition," Biol. Cybern. **116**, in press (2022).
[186] G. B. Ermentrout and D. H. Terman, *Mathematical Foundations of Neuroscience* (Springer, 2010).
[187] J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina, "Higher-order interactions stabilize dynamics in competitive network models," Nature **548**, 210 (2017).
[188] J. M. Levine, J. Bascompte, P. B. Adler, and S. Allesina, "Beyond pairwise mechanisms of species coexistence in complex communities," Nature **546**, 56 (2017).
[189] R. Lambiotte, M. Rosvall, and I. Scholtes, "From networks to optimal higher-order models of complex systems," Nat. Phys. **15**, 313 (2019).
[190] 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).
[191] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. F. de Arruda, B. Franceschiello, I. Iacopini, and S. Kéfi, "The physics of higher-order interactions in complex systems," Nat. Phys. **17**, 1093 (2021).
[192] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer, 2004).
[193] G. St-Onge, *Processus de contagion sur réseaux complexes au-delà des interactions dyadiques*, Ph.D. thesis, Université Laval (2022).
[194] G. St-Onge, *Dynamique de propagation sur réseaux aléatoires: Caractérisation de la transition de phase*, Master's thesis, Université Laval (2017).
[195] J. D. Cowan, "Stochastic neurodynamics," in *Adv. Neural Inf. Process. Syst.*, Vol. 3 (1990) p. 62.
[196] R. J. Glauber, "Time-Dependent Statistics of the Ising Model," J. Math. Phys. **4**, 294 (1963).
[197] C. Murphy, V. Thibeault, A. Allard, and P. Desrosiers, "Duality between predictability and reconstructability in complex systems," arXiv:2206.04000 (2022).
[198] R. H. R. Hahnloser and H. S. Seung, "Permitted and forbidden sets in symmetric threshold-linear networks," in *Adv. Neural Inf. Process. Syst.* (2000).
[199] C. Parmelee, J. L. Alvarez, C. Curto, and K. Morrison, "Sequential attractors in combinatorial threshold-linear networks," SIAM J. Appl. Dyn. Syst. **21**, 1597 (2022).
[200] F. Morone, G. D. Ferraro, and H. A. Makse, "The k-core as a predictor of structural collapse in mutualistic ecosystems," Nat. Phys. **15**, 95 (2019).
[201] G. Li, "A lumping analysis in mono- or/and bimolecular reaction systems," Chem. Eng. Sci. **39**, 1261 (1984).
[202] P. W. Milonni and J. H. Eberly, *Lasers* (Wiley, 1988).

[203] A. J. Lotka, "Contribution to the theory of periodic reactions," J. Phys. Chem. **14**, 271 (1910).
[204] V. Volterra, "Fluctuations in the abundance of a species considered mathematically," Nature **118**, 558 (1926).
[205] H. Sanhedrai, J. Gao, A. Bashan, M. Schwartz, S. Havlin, and B. Barzel, "Reviving a failed network through microscopic interventions," Nat. Phys. **18**, 338 (2022).
[206] W. C. Allee and E. S. Bowen, "Studies in animal aggregations: Mass protection against colloidal silver among goldfishes," J. Exp. Zool. **61**, 185 (1932).
[207] J. Gao, B. Barzel, and A.-L. Barabási, "Universal resilience patterns in complex networks," Nature **530**, 307 (2016).
[208] Y. Kuramoto, "Self-entrainment of a population of coupled non-linear oscillators," in *International Symposium on Mathematical Problems in Theoretical Physics* (1975) p. 420.
[209] H. Sakaguchi and Y. Kuramoto, "A soluble active rotator model showing phase transitions via mutual entrainment," Prog. Theor. Phys. **76**, 576 (1986).
[210] B. Pietras and A. Daffertshofer, "Network dynamics of coupled oscillators and phase reduction techniques," Phys. Rep. **819**, 1 (2019).
[211] K. Wiesenfeld, P. Colet, and S. H. Strogatz, "Synchronization transitions in a disordered Josephson series array," Phys. Rev. Lett. **76**, 404 (1996).
[212] M. H. Matheny, J. Emenheiser, W. Fon, A. Chapman, A. Salova, M. Rohden, J. Li, M. Hudoba De Badyn, M. Pósfai, L. Duenas-Osorio, M. Mesbahi, J. P. Crutchfield, M. C. Cross, R. M. D'Souza, and M. L. Roukes, "Exotic states in a simple network of nanoelectromechanical oscillators," Science **363**, 1057 (2019).
[213] E. M. Izhikevich, *Dynamical Systems in Neuroscience* (MIT Press, 2007).
[214] A. T. Winfree, "Biological rhythms and the behavior of populations of coupled oscillators," J. Theoret. Biol. **16**, 15 (1967).
[215] G. B. Ermentrout and N. Kopell, "Parabolic bursting in an excitable system coupled with a slow oscillation," SIAM J. Appl. Math. **46**, 233 (1986).
[216] K. I. Funahashi and Y. Nakamura, "Approximation of dynamical systems by continuous time recurrent neural networks," Neural Netw. **6**, 801 (1993).
[217] R. D. Beer, "The dynamics of adaptive behavior: A research program," Robot. Auton. Syst. **20**, 257 (1997).
[218] M. A. Cohen and S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," IEEE Trans. Syst. Man Cybern. **SMC-13**, 815 (1983).
[219] S. Grossberg, "Nonlinear neural networks: Principles, mechanisms, and architectures," Neural Netw. **1**, 17 (1988).
[220] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," Proc. Natl. Acad. Sci. U.S.A. **81**, 3088 (1984).
[221] M. W. Hirsch, "Convergent activation dynamics in continuous time networks," Neural Netw. **2**, 331 (1989).
[222] R. Kim, Y. Li, and T. J. Sejnowski, "Simple framework for constructing functional spiking recurrent neural networks," Proc. Natl. Acad. Sci. U.S.A. **116**, 22811 (2019).
[223] R. Kim and T. J. Sejnowski, "Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks," Nat. Neurosci. **24**, 129 (2021).
[224] J. Hanson and M. Raginsky, "Universal simulation of stable dynamical systems by recurrent neural nets," Proc. Mach. Learn. Res. **120**, 1 (2020).
[225] L. Qi and Z. Luo, *Tensor analysis* (SIAM, 2017).
[226] C. Berge, *Hypergraphs* (North-Holland, 1989).
[227] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen, "Directed hypergraphs and applications," Discret. Appl. Math. **42**, 177 (1993).
[228] G. Ferraz de Arruda, M. Tizzani, and Y. Moreno, "Phase transitions and stability of dynamical processes on hypergraphs," Commun. Phys. **4**, 24 (2021).
[229] R. Mulas, C. Kuehn, and J. Jost, "Coupled dynamics on hypergraphs: Master stability of steady states and synchronization," Phys. Rev. E **101**, 062313 (2020).
[230] P. Ashwin and A. Rodrigues, "Hopf normal form with SN symmetry and reduction to systems of nonlinearly coupled phase oscillators," Physica **325**, 14 (2016).
[231] I. León and D. Pazó, "Phase reduction beyond the first order : The case of the mean-field complex Ginzburg-Landau equation," Phys. Rev. E **100**, 012211 (2019).
[232] E. Nijholt, J. L. Ocampo-Espindola, D. Eroglu, I. Z. Kiss, and T. Pereira, "Emergent hypernetworks in weakly coupled oscillators," Nat. Commun. **13**, 4849 (2022).
[233] E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, "Spectral dimension reduction of complex dynamical networks," Phys. Rev. X **9**, 011042 (2019).
[234] P. Kundu, H. Kori, and N. Masuda, "Accuracy of a one-dimensional reduction of dynamical systems on networks," Phys. Rev. E **105**, 024305 (2022).
[235] C. Tu, J. Grilli, F. Schuessler, and S. Suweis, "Collapse of resilience patterns in generalized Lotka-Volterra dynamics and beyond," Phys. Rev. E **95**, 062307 (2017).
[236] J. Jiang, Z.-G. Huang, T. P. Seager, W. Lin, C. Grebogi, A. Hastings, and Y.-C. Lai, "Predicting tipping points in mutualistic networks through dimension reduction." Proc. Natl. Acad. Sci. U.S.A. **115**, E639 (2018).
[237] M. Vegué, V. Thibeault, P. Desrosiers, and A. Allard, "Dimension reduction of dynamics on modular and heterogeneous directed networks," PNAS Nexus , pgad150 (2023).
[238] D. Hadjiabadi, M. Lovett-Barron, I. G. Raikov, F. T. Sparks, Z. Liao, S. C. Baraban, J. Leskovec, A. Losonczy, K. Deisseroth, and I. Soltesz, "Maximally selective single-cell target for circuit control in epilepsy models," Neuron **109**, 2556

(2021).
[239] R. Lim, J. J. T. Cabatbat, T. L. P. Martin, H. Kim, S. Kim, J. Sung, C. M. Ghim, and P. J. Kim, "Large-scale metabolic interaction network of the mouse and human gut microbiota," Sci. Data **7**, 1 (2020).
[240] C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," IEEE Trans. Pattern Anal. Mach. Intell. **32**, 45 (2010).
[241] P. Städter, Y. Schälte, L. Schmiester, J. Hasenauer, and P. L. Stapor, "Benchmarking of numerical integration methods for ODE models of biological systems," Sci. Rep. **11**, 2696 (2021).
[242] B. G. Fenyves, G. S. Szilágyi, Z. Vassy, C. Sőti, and P. Csermely, "Synaptic polarity and sign-balance prediction using gene expression data in the Caenorhabditis elegans chemical synapse neuronal connectome network," PLOS Comput. Biol. **16**, e1007974 (2020).
[243] L. K. Scheffer *et al.*, "A connectome and analysis of the adult Drosophila central brain," eLife **9**, 1 (2020).
[244] K. Ryan, Z. Lu, and I. A. Meinertzhagen, "The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling," eLife **5**, 1 (2016).
[245] C. Verasztó, S. Jasek, M. Gühmann, R. Shahidi, N. Ueda, J. D. Beard, S. Mendes, K. Heinz, L. A. Bezares-Calderón, E. Williams, and G. Jékely, "Whole-animal connectome and cell-type complement of the three-segmented Platynereis dumerilii larva," bioRxiv (2020).
[246] S. Jasek, C. Verasztó, E. Brodrick, R. Shahidi, T. Kazimiers, and G. Jékely, "Desmosomal connectomics of all somatic muscles in an annelid larva," eLife **11**, e71231 (2022).
[247] S. W. Oh and al., "A mesoscale connectome of the mouse brain," Nature **508**, 207 (2014).
[248] L. Coletta, M. Pagani, J. D. Whitesell, J. A. Harris, B. Bernhardt, and A. Gozzi, "Network structure of the mouse brain connectome with voxel resolution," Sci. Adv. **6**, eabb7187 (2020).
[249] J. Wachs, M. Fazekas, and J. Kertész, "Corruption risk in contracting markets: a network science perspective," Int. J. Data Sci. Anal. **12**, 45 (2021).
[250] S. Ranganathan, M. Kivelä, and J. Kanniainen, "Dynamics of investor spanning trees around dot-com bubble," PLOS ONE **13**, e0198807 (2018).

Supplementary Table 1: Real networks and their effective ranks

The following table lists all the real networks that were used in the paper, where all the names have a clickable link to the source of the data, along with their number of vertices N, their rank, and their effective ranks. We have shortened the names of six celegans networks by replacing "2019_hermaphrodite_" by "...".

Name	N	rank	srank	nrank	elbow	energy	thrank	shrank	erank
7th_graders	29	29	1.7	4.6	4	10	2	3	18.8
AT_2008	2271	842	51.6	170.0	376	410	82	108	687.5
CY_2015	335	108	14.6	33.3	108	54	6	8	90.1
EE_2010	1260	312	6.1	29.9	60	114	20	34	227.5
PT_2009	1257	560	19.0	84.4	128	309	16	36	471.7
SI_2016	1792	790	11.2	65.8	136	281	80	108	566.3
adjnoun	112	109	4.9	17.0	10	42	10	15	75.2
advogato	6541	5550	15.7	169.8	302	993	905	1053	3180.3
ambassador_1985_1989	16	11	1.9	3.6	1	5	1	1	8.4
ambassador_1990_1994	16	14	2.1	4.1	3	6	2	2	10.6
ambassador_1995_1999	16	15	1.6	3.4	1	5	3	3	10.1
ambassador_2000	16	15	1.4	3.1	1	5	1	1	9.9
ambassador_2001	16	15	1.4	3.1	1	5	1	1	9.9
ambassador_2002	16	14	1.4	2.9	1	5	1	2	9.0
ambassador_OPERATION_ID	16	15	1.4	3.1	1	5	1	1	9.9
ambassador_TIE_EXTINGUISH	16	15	1.4	3.1	1	5	1	1	9.9
ambassador_TIE_YEAR	16	15	1.4	3.1	1	5	1	1	9.9
anybeat	12645	4152	9.4	111.1	184	788	583	711	2487.6
arxiv_collab_astro-ph-1999	16706	15668	73.4	558.9	1120	2315	3643	4097	7040.2
arxiv_collab_cond-mat-1999	16726	15523	156.7	872.5	1537	2649	3370	3791	7801.7
arxiv_collab_hep-th-1999	8361	7101	44.7	340.9	485	1539	1507	1722	3926.6
baseball_player-player	72	72	1.1	3.2	4	2	7	8	28.3
baseball_user-provider	84	24	3.8	7.7	8	12	2	2	19.3
bible_nouns	1773	1757	2.7	23.0	59	68	351	404	636.0
bison	26	26	1.8	4.3	5	7	4	4	15.3
bitcoin_alpha	3783	1982	13.5	107.2	152	538	304	374	1278.2
bitcoin_trust	5881	2785	15.8	134.9	197	721	411	504	1778.2
blumenau_drug	75	56	5.9	12.7	17	17	13	15	34.6
board_directors_net1m_2002-05-01	1013	1012	10.1	69.7	84	374	144	155	733.6
board_directors_net1m_2002-12-01	1076	1076	23.1	110.3	185	406	150	163	781.3
board_directors_net1m_2003-01-01	1098	1098	21.3	105.9	187	400	153	164	792.8
board_directors_net1m_2003-12-01	1142	1142	22.1	109.8	190	418	156	172	826.5
board_directors_net1m_2004-01-01	1145	1145	22.1	110.2	191	419	157	173	829.4
board_directors_net1m_2004-12-01	1194	1194	22.9	114.9	144	434	164	182	866.5
board_directors_net1m_2005-01-01	1198	1198	23.1	115.7	144	434	167	185	870.1
board_directors_net1m_2005-12-01	1310	1310	23.3	125.6	226	544	184	202	972.4
board_directors_net1m_2006-01-01	1316	1316	23.4	126.0	227	547	186	202	977.3
board_directors_net1m_2006-12-01	1471	1469	40.8	177.5	273	608	219	248	1094.7
board_directors_net1m_2007-01-01	1499	1497	41.6	180.7	281	619	226	253	1114.4
board_directors_net1m_2007-12-01	1521	1517	38.0	176.4	202	630	230	260	1143.4
board_directors_net1m_2008-01-01	1518	1513	43.7	190.3	309	638	231	264	1145.5
board_directors_net1m_2008-12-01	1519	1515	68.1	242.2	316	673	238	268	1159.9
board_directors_net1m_2009-01-01	1526	1523	68.2	242.8	313	680	235	265	1166.3
board_directors_net1m_2009-12-01	1483	1480	34.9	169.9	252	656	220	249	1131.8

board_directors_net1m_2010-01-01	1473	1470	34.6	168.6	254	649	220	249	1123.7
board_directors_net1m_2010-12-01	1444	1442	59.9	221.6	288	649	223	248	1106.4
board_directors_net1m_2011-01-01	1442	1440	60.3	221.7	289	642	224	249	1103.3
budapest_connectome_all_1m	1015	1015	3.1	21.1	28	73	191	232	422.5
budapest_connectome_all_200k	1015	1015	3.6	27.1	36	140	170	214	493.7
budapest_connectome_all_20k	1015	1015	5.3	40.0	40	225	139	185	577.6
budapest_connectome_female_1m	1015	1015	3.1	21.8	31	83	190	231	430.8
budapest_connectome_female_200k	1015	1015	3.7	27.7	34	142	168	210	494.1
budapest_connectome_female_20k	1015	1015	5.6	41.4	46	225	148	193	574.5
budapest_connectome_male_1m	1015	1015	3.1	20.7	26	66	215	250	395.5
budapest_connectome_male_200k	1015	1015	3.6	26.2	29	119	201	240	461.4
budapest_connectome_male_20k	1015	1015	5.7	41.0	45	215	139	186	566.6
caida_as_20040105	16301	4679	16.8	169.2	208	1169	521	671	3096.8
caida_as_20040202	16493	4669	17.0	170.1	197	1180	521	665	3094.6
caida_as_20040301	16655	4771	17.1	173.4	223	1216	543	676	3173.6
caida_as_20040405	16874	4829	17.0	173.1	207	1220	556	690	3204.6
caida_as_20040503	17160	4904	17.3	177.0	244	1240	569	721	3256.1
caida_as_20040607	17306	4924	17.4	177.1	268	1230	580	717	3255.4
caida_as_20040705	17509	4991	17.7	179.9	250	1248	584	736	3303.2
caida_as_20040802	17655	5056	18.1	183.4	257	1271	592	737	3348.7
caida_as_20040906	17848	5084	17.6	180.3	256	1255	594	747	3352.3
caida_as_20041004	18100	5213	17.7	182.7	241	1282	594	752	3434.2
caida_as_20041101	18278	5190	17.8	183.4	228	1302	596	761	3436.6
caida_as_20041206	18501	5234	18.0	184.7	239	1287	608	772	3447.2
caida_as_20050103	18740	5312	18.4	189.2	263	1329	620	787	3519.4
caida_as_20050207	18911	5323	18.6	191.2	242	1351	619	785	3539.1
caida_as_20050307	19090	5334	18.4	190.0	272	1349	623	784	3547.1
caida_as_20050404	19267	5415	18.6	192.0	224	1356	638	799	3589.6
caida_as_20050502	19489	5516	18.5	192.8	277	1376	649	818	3652.6
caida_as_20050606	19720	5542	18.5	193.6	278	1387	632	800	3668.5
caida_as_20050704	19846	5538	18.4	192.7	262	1382	638	802	3663.3
caida_as_20070917	8020	2595	11.4	95.9	129	494	255	323	1605.8
cattle	28	25	2.2	4.8	6	7	5	6	14.3
celegans_...chemical	453	300	10.6	35.1	47	69	56	69	173.1
celegans_...chemical_corrected	454	300	10.6	35.1	47	69	56	69	173.1
celegans_...chemical_synapse	468	271	9.8	33.1	33	71	52	61	158.7
celegans_...gap_junction	468	460	2.9	9.7	19	4	100	117	119.4
celegans_...gap_junction_corrected	469	460	2.7	8.7	19	4	98	116	106.6
celegans_...gap_junction_synapse	466	303	5.9	23.8	28	58	51	61	159.2
celegans_2019_male_chemical	575	375	9.9	35.7	48	77	83	95	197.3
celegans_2019_male_chemical_corrected	575	375	9.9	35.7	48	77	83	95	197.3
celegans_2019_male_chemical_synapse	514	296	8.9	30.0	41	61	68	80	149.1
celegans_2019_male_gap_junction	585	536	3.1	13.7	12	21	120	146	173.4
celegans_2019_male_gap_junction_corrected	586	537	3.1	13.7	12	21	122	148	173.3
celegans_2019_male_gap_junction_synapse	514	298	5.2	18.9	33	35	69	78	120.4
celegans_interactomes_BPmaps	537	318	34.2	90.0	59	180	9	19	273.6
celegans_interactomes_Genetic	759	619	21.0	86.1	79	253	55	78	459.2
celegans_interactomes_IntegratedNetwork	6176	5102	2.7	49.4	108	547	1038	1205	2415.5
celegans_interactomes_Interolog	2724	2052	5.7	59.4	67	439	243	314	1264.0
celegans_interactomes_LCI	431	244	8.2	35.8	20	124	16	20	199.0
celegans_interactomes_Microarray	2436	2180	2.1	27.7	52	256	482	565	1021.3
celegans_interactomes_Phenotypes	912	912	5.0	35.8	46	147	197	229	474.8
celegans_interactomes_WI8	2528	1523	40.6	193.1	146	675	113	156	1184.9
celegans_interactomes_wi2004	1237	702	24.9	104.5	80	330	44	61	557.0
celegans_interactomes_wi2007	1496	906	35.5	143.4	108	450	62	85	732.3
celegans_metabolic	453	450	3.0	13.2	25	15	54	66	161.8

celegans_signed	297	292	8.7	29.6	39	63	70	80	149.1
celegansneural	297	248	3.1	13.2	16	35	67	74	103.3
ceo_club	40	30	4.5	9.8	4	18	2	2	25.5
chess	7301	6383	23.6	252.0	334	1752	1021	1261	3900.2
chicago_road	12982	12666	2183.5	4427.1	2258	6277	9	332	10134.0
cintestinalis	213	203	3.7	12.8	19	23	52	60	81.6
college_freshmen	32	31	3.3	7.7	5	15	2	3	22.9
collins_yeast	1622	1558	5.1	49.0	54	296	231	284	956.3
contact	274	82	1.3	4.5	1	13	23	26	32.2
contiguous_usa	49	49	7.6	16.0	9	26	2	3	39.3
copenhagen_bt	692	692	20.6	68.9	119	137	192	213	324.9
copenhagen_calls	536	420	6.0	19.7	38	30	113	130	118.7
copenhagen_fb_friends	800	799	12.8	71.4	60	282	79	112	537.4
copenhagen_sms	568	502	3.7	12.7	25	13	124	137	117.3
crime	1380	900	94.6	254.5	140	516	10	30	777.3
cs_department	61	61	3.7	10.4	6	20	8	11	39.6
dblp_cite	12590	3137	42.8	290.7	214	1495	168	251	2501.0
dnc	2029	627	2.3	15.5	30	53	153	176	280.2
dolphins	62	60	6.1	14.9	8	26	3	6	43.8
drosophila	21733	21687	11.6	173.5	724	1062	4868	5543	6702.8
dutch_criticism	35	35	3.6	8.4	8	15	3	3	25.0
dutch_school_klas12b-net-1	26	24	3.2	7.0	3	12	2	2	18.7
dutch_school_klas12b-net-2	26	25	2.2	5.3	3	10	2	3	17.6
dutch_school_klas12b-net-3	26	22	1.6	4.1	3	8	2	3	14.8
dutch_school_klas12b-net-3m	26	22	1.6	4.1	3	8	2	3	14.8
dutch_school_klas12b-net-4	26	24	2.1	5.2	4	10	1	2	17.2
dutch_school_klas12b-net-4m	26	24	2.1	5.2	4	10	1	2	17.2
dutch_school_klas12b-primary	26	23	1.8	4.7	2	11	1	2	17.4
ecoli_transcription_v1_0	424	115	7.4	24.0	18	62	8	12	95.8
ecoli_transcription_v1_1	423	115	7.3	23.9	18	62	8	12	95.7
edit_wikibooks_aa	115	72	6.3	17.3	8	40	4	6	60.0
edit_wikibooks_az	6596	772	6.4	24.4	40	31	176	192	248.2
edit_wikibooks_be	1160	360	4.4	13.4	26	9	58	70	112.5
edit_wikibooks_bs	1135	362	5.6	15.3	22	10	46	56	112.7
edit_wikibooks_ca	8052	1541	9.7	39.4	106	65	448	484	348.7
edit_wikibooks_cy	781	326	5.0	18.6	34	26	62	72	159.1
edit_wikibooks_da	4567	884	3.8	13.0	30	11	196	218	153.9
edit_wikibooks_eu	1197	361	3.4	9.1	20	9	74	80	64.6
edit_wikibooks_fa	16002	2396	5.7	26.7	96	32	530	598	509.6
edit_wikibooks_fy	739	288	3.2	10.4	18	7	36	42	102.0
edit_wikibooks_gl	3466	494	2.8	9.4	30	10	96	112	101.2
edit_wikibooks_hr	3771	522	6.4	19.3	54	23	114	124	138.1
edit_wikibooks_ie	705	258	3.6	11.1	14	6	32	40	96.9
edit_wikibooks_ka	5267	494	4.1	12.8	38	16	102	112	101.8
edit_wikibooks_ky	647	270	2.8	9.3	12	6	38	42	97.9
edit_wikibooks_la	1014	377	3.4	10.3	18	9	64	74	93.1
edit_wikibooks_my	98	50	3.7	9.4	8	16	8	8	35.0
edit_wikibooks_na	300	100	2.9	7.5	6	6	16	16	44.1
edit_wikibooks_no	4821	1273	4.2	21.9	66	51	378	406	319.0
edit_wikibooks_oc	875	304	3.5	11.9	20	10	44	52	110.8
edit_wikibooks_si	3427	429	6.8	21.5	52	27	88	104	144.3
edit_wikibooks_sr	3308	627	6.8	27.2	56	48	156	170	229.5
edit_wikibooks_tg	960	274	4.8	18.5	30	32	40	48	146.0
edit_wikibooks_tt	1671	342	5.4	18.7	24	21	54	58	148.9
edit_wikibooks_ur	1584	433	12.7	34.4	62	32	80	88	177.8
edit_wikibooks_xh	167	74	2.2	5.1	6	2	6	8	31.2

edit_wikibooks_zh	14125	3699	13.2	76.9	280	179	854	950	1080.8
edit_wikibooks_zu	271	122	3.7	8.8	12	5	12	14	49.7
edit_wikinews_bg	4457	603	3.0	4.7	12	4	122	136	20.1
edit_wikinews_el	10982	563	2.2	3.0	8	2	132	144	7.6
edit_wikinews_hu	5541	612	2.8	6.6	20	4	134	158	61.0
edit_wikinews_no	4523	816	3.2	13.8	42	16	170	194	206.6
edit_wikinews_sv	9105	940	2.3	5.4	14	3	174	206	61.8
edit_wikiquote_als	111	68	4.1	11.4	10	27	4	8	48.9
edit_wikiquote_et	1822	580	6.0	21.5	42	21	128	140	198.7
edit_wikiquote_id	4414	1159	10.2	42.1	104	58	196	238	439.1
edit_wikiquote_ml	3834	811	4.2	14.9	52	14	182	206	184.5
edit_wikiquote_te	3795	529	3.1	11.1	30	12	120	138	119.2
edit_wikiquote_zh	11164	3442	7.5	45.2	158	76	726	808	978.7
edit_wikiquote_zh_min_nan	550	208	5.3	17.3	28	24	28	34	109.2
edit_wiktionary_be	7426	594	4.5	15.8	48	22	136	156	151.3
edit_wiktionary_fo	2064	464	3.7	12.9	34	15	98	108	124.4
edit_wiktionary_mi	2262	398	3.7	13.2	44	17	88	96	124.7
edit_wiktionary_roa_rup	2137	334	6.2	16.6	32	15	70	86	101.8
edit_wiktionary_st	2312	366	3.9	13.3	26	17	66	76	114.9
edit_wiktionary_vec	3366	349	3.4	9.5	26	9	78	92	72.6
edit_wiktionary_zu	1624	418	3.3	12.6	26	14	76	86	135.4
ego_social_facebook_0	333	327	3.7	21.6	19	87	52	66	203.4
ego_social_facebook_107	1034	1034	3.5	33.8	32	252	115	161	617.3
ego_social_facebook_1684	786	783	5.7	40.9	26	221	92	127	488.5
ego_social_facebook_1912	747	745	2.3	19.8	23	135	73	106	422.2
ego_social_facebook_3437	534	531	6.4	36.9	31	153	60	84	334.8
ego_social_facebook_348	224	223	2.6	14.4	13	61	26	36	137.6
ego_social_facebook_3980	52	51	3.3	9.4	4	19	7	9	36.0
ego_social_facebook_414	150	149	2.7	11.5	11	37	15	20	87.9
ego_social_facebook_686	168	167	2.8	13.6	11	50	18	23	105.3
ego_social_facebook_698	61	60	2.5	8.1	6	19	6	6	39.3
ego_social_facebook_combined	4039	3955	6.7	87.2	109	859	515	691	2304.8
ego_social_gplus_100129275726588145876	1650	1624	2.7	28.6	43	217	312	384	774.1
ego_social_gplus_100329698645326486178	2213	2183	5.3	60.5	64	473	453	533	1172.2
ego_social_gplus_100466178325794757407	344	331	3.1	18.3	18	78	44	60	189.0
ego_social_gplus_100500197140377336562	638	622	3.9	26.8	31	128	113	138	328.5
ego_social_gplus_100518419853963396365	326	326	2.3	15.4	13	82	35	50	189.2
elec	7118	2382	9.7	100.8	78	721	333	431	1565.2
elite	44	40	5.2	11.1	6	17	2	4	29.2
email_company	167	138	3.9	11.1	13	18	36	42	52.6
escorts	16730	9978	48.4	368.6	630	1543	1570	1894	5477.5
eu_airlines	450	326	2.6	15.4	21	61	69	81	168.3
euroroad	1174	1125	176.2	386.5	136	620	2	18	943.8
faa_routes	1226	858	63.6	198.8	129	490	51	78	727.2
facebook_friends	362	342	6.7	29.7	26	94	44	59	214.1
facebook_organizations_L1	5793	3412	14.1	124.8	172	677	630	747	1935.9
facebook_organizations_L2	5524	5524	30.8	273.8	275	1732	646	877	3558.3
facebook_organizations_M1	1429	1428	8.1	67.4	85	386	217	277	858.5
facebook_organizations_M2	3862	3862	33.8	247.6	186	1320	353	514	2573.8
facebook_organizations_S1	320	320	5.4	27.8	15	102	38	52	206.4
facebook_organizations_S2	165	155	5.2	19.7	16	54	16	21	102.5
fao_trade	214	159	2.0	4.6	8	5	59	62	21.5
fediverse	4860	806	1.7	16.1	18	122	32	44	483.3
financial_institution07-Apr-1999	31	31	3.2	7.7	4	16	2	2	23.5
florentine_families	16	15	3.2	5.5	6	7	1	3	11.2
foldoc	13356	12919	65.1	636.3	511	4244	1747	2198	8449.5

foodweb_baywet	128	98	2.7	4.2	11	5	41	41	7.7
foodweb_little_rock	183	45	2.6	6.6	5	10	8	11	25.9
football	115	115	10.5	26.0	13	46	10	11	81.2
football_tsevans	115	115	10.5	26.0	13	46	10	11	81.2
foursquare_NYC_restaurant_checkin	4936	4020	17.5	136.0	184	662	620	784	2091.8
foursquare_NYC_restaurant_tips	6410	4980	64.0	430.5	406	2116	400	572	3804.5
fresh_webs_AkatoreA	85	35	2.4	6.6	5	13	3	5	24.9
fresh_webs_AkatoreB	58	17	2.5	5.2	3	9	1	1	13.6
fresh_webs_Berwick	79	32	2.2	6.1	2	13	2	3	22.7
fresh_webs_Blackrock	87	33	2.3	6.7	2	16	1	1	25.5
fresh_webs_Broad	95	39	1.8	5.6	1	15	1	2	27.7
fresh_webs_Canton	110	43	2.0	6.4	3	17	2	3	30.7
fresh_webs_Catlins	49	18	2.0	4.6	4	8	1	2	13.8
fresh_webs_Coweeta1	58	20	2.5	5.7	2	10	1	1	16.2
fresh_webs_Coweeta17	71	26	2.4	6.1	3	12	2	2	20.2
fresh_webs_DempstersAu	86	36	2.0	5.9	3	15	2	2	26.2
fresh_webs_DempstersSp	97	39	1.9	6.0	4	15	2	4	27.5
fresh_webs_DempstersSu	107	45	1.9	6.3	3	18	2	3	32.6
fresh_webs_German	86	32	2.3	6.3	4	14	2	3	23.5
fresh_webs_Healy	96	40	2.0	6.2	2	16	2	3	28.6
fresh_webs_Kyeburn	98	38	1.9	5.8	3	15	1	3	27.0
fresh_webs_LilKyeburn	78	27	2.2	5.9	3	12	2	3	20.6
fresh_webs_Martins	105	41	3.3	9.0	6	18	3	4	31.0
fresh_webs_Narrowdale	71	26	2.3	5.6	4	10	2	4	18.6
fresh_webs_NorthCol	78	29	2.6	6.5	6	12	3	4	21.2
fresh_webs_Powder	78	35	2.2	6.3	3	13	2	4	24.9
fresh_webs_Stony	113	46	1.7	5.8	2	17	2	3	32.3
fresh_webs_SuttonAu	83	31	1.9	5.6	1	14	1	1	23.3
fresh_webs_SuttonSp	79	24	1.7	4.7	1	11	1	1	18.1
fresh_webs_SuttonSu	94	17	1.5	3.6	1	8	1	1	12.7
fresh_webs_Troy	78	28	2.9	7.2	2	14	1	2	22.3
fresh_webs_Venlaw	69	27	2.3	6.1	2	12	1	2	20.8
freshmen_t0	34	11	1.0	1.1	1	1	1	1	1.9
freshmen_t2	34	34	1.1	2.1	1	1	1	2	11.0
freshmen_t3	34	34	1.0	2.0	1	1	1	2	10.4
freshmen_t5	34	34	1.0	1.9	2	1	2	2	8.6
freshmen_t6	34	30	1.1	1.9	2	1	2	2	8.4
fully_connected_layer_cnn_00100	820	299	1.2	1.7	4	2	124	125	3.1
fully_connected_layer_cnn_00200	980	10	1.1	1.4	3	1	3	3	2.2
fully_connected_layer_cnn_00300	900	128	5.0	16.4	124	39	12	16	81.1
fully_connected_layer_cnn_00400	1028	256	9.3	27.3	129	47	65	77	126.8
fully_connected_layer_cnn_00500	892	380	20.6	78.4	51	237	4	9	336.4
fully_connected_layer_cnn_00600	1076	308	30.2	85.1	308	199	8	14	273.8
fully_connected_layer_cnn_00700	1028	256	28.2	69.6	256	135	13	22	206.0
fully_connected_layer_cnn_00800	628	116	11.4	29.5	116	60	7	10	92.2
fully_connected_layer_cnn_00900	396	128	1.2	2.1	7	2	40	44	5.9
fully_connected_layer_cnn_01000	820	308	19.8	56.6	97	117	28	36	221.1
game_thrones	107	98	5.0	12.6	22	18	25	28	46.0
genetic_multiplex_Arabidopsis	6980	2509	15.3	138.7	127	873	277	362	1798.5
genetic_multiplex_Bos_Multiplex_Genetic	325	119	13.3	35.3	15	77	2	5	105.9
genetic_multiplex_Candida	367	46	2.0	6.5	8	17	5	7	32.9
genetic_multiplex_Celegans	3879	1078	12.1	82.0	73	398	107	142	792.5
genetic_multiplex_DanioRerio	155	74	6.4	18.6	10	46	5	8	63.8
genetic_multiplex_Drosophila	8215	4877	45.0	322.4	313	1538	678	865	3259.7
genetic_multiplex_Homo	18222	8436	14.1	199.8	338	1757	1492	1794	4746.9
genetic_multiplex_HumanHerpes4	216	25	1.8	4.8	1	10	1	1	18.3

genetic_multiplex_Mus	7747	2552	18.6	154.6	176	915	272	363	1838.4
genetic_multiplex_Oryctolagus	144	26	2.0	5.4	3	12	2	2	20.1
genetic_multiplex_Plasmodium	1203	575	27.1	100.3	80	265	49	70	455.8
genetic_multiplex_Rattus	2640	752	6.1	49.1	43	312	43	67	577.6
genetic_multiplex_Sacchcere	6570	5116	12.1	149.6	175	1203	1010	1210	2866.0
genetic_multiplex_Sacchpomb	4092	1734	9.0	80.2	89	450	318	384	1060.6
genetic_multiplex_Xenopus	461	195	17.1	49.0	30	109	8	14	165.5
genetic_multiplex_YeastLandscape	4458	4457	1.4	13.6	54	5	679	857	1394.9
gnutella_04	10879	4860	168.3	779.4	213	2640	38	89	4086.0
gnutella_06	8717	3680	84.4	482.6	120	2061	14	43	3139.1
gnutella_08	6301	2195	36.2	245.1	93	1282	31	48	1896.9
gnutella_09	8114	2759	42.1	297.9	104	1629	33	50	2396.1
google	15763	9655	6.5	119.0	303	1110	1418	1752	5456.2
gut	838	735	2.5	7.4	35	7	197	203	76.2
hens	32	31	1.3	3.4	2	6	2	3	17.5
high_school_proximity	327	327	6.8	30.1	16	104	16	28	209.5
high_tech_company	21	21	1.7	3.9	1	7	1	3	12.9
highschool	70	67	6.5	16.1	12	29	4	6	49.4
households_04-Sep-1998	111	111	2.0	8.9	15	36	9	10	72.5
households_09-Jan-2002	695	695	2.3	15.2	50	23	48	52	327.8
interactome_figeys	2239	333	6.8	33.9	34	113	42	55	236.2
interactome_pdz	212	110	14.4	34.4	26	60	2	10	94.3
interactome_stelzl	1706	941	20.2	104.2	113	379	99	131	713.9
interactome_vidal	3133	2356	52.9	267.7	240	968	205	284	1743.2
interactome_yeast	1870	1263	82.2	267.9	169	657	59	98	1045.1
jazz_collab	198	198	3.4	15.0	13	45	25	32	114.9
jdk	6434	2180	1.8	12.4	56	14	432	511	490.0
jung	6120	2035	2.0	13.8	67	21	424	485	475.5
kangaroo	17	17	1.4	2.5	3	3	5	5	6.7
karate_77	34	24	3.5	7.2	6	11	3	3	18.6
karate_78	34	24	3.4	7.2	6	11	3	3	18.4
kegg_metabolic_aae	926	854	9.1	59.0	73	290	85	111	591.1
kegg_metabolic_afu	909	820	9.4	60.1	63	309	66	82	586.4
kegg_metabolic_ana	1362	1245	8.4	65.6	89	403	113	147	853.6
kegg_metabolic_ape	805	726	9.2	55.9	57	271	62	80	515.4
kegg_metabolic_atc	1589	1447	8.6	70.9	94	466	119	156	989.4
kegg_metabolic_ath	1561	1409	8.8	70.7	94	447	108	146	960.6
kegg_metabolic_atu	1593	1451	8.6	71.0	94	467	119	158	992.2
kegg_metabolic_bas	664	618	9.8	54.3	55	231	55	73	436.6
kegg_metabolic_bbu	505	465	8.3	44.1	42	183	37	51	335.5
kegg_metabolic_bha	1267	1169	9.0	67.2	79	386	108	141	804.4
kegg_metabolic_bja	1746	1583	8.6	73.1	92	492	129	171	1074.7
kegg_metabolic_blo	854	795	8.9	56.7	79	270	89	112	553.7
kegg_metabolic_bme	1574	1426	8.6	70.8	88	462	119	152	977.4
kegg_metabolic_bms	1405	1286	9.0	69.7	83	427	108	140	887.2
kegg_metabolic_bsu	1380	1275	8.8	68.6	85	414	109	148	871.9
kegg_metabolic_buc	638	601	10.0	54.0	48	222	53	70	419.8
kegg_metabolic_cac	1081	1006	8.8	62.6	80	331	107	140	691.5
kegg_metabolic_ccr	1421	1297	8.6	68.1	92	432	108	138	896.9
kegg_metabolic_cef	1166	1092	8.4	62.3	79	351	106	142	744.5
kegg_metabolic_cel	1307	1187	8.9	66.3	87	383	102	135	815.1
kegg_metabolic_cgl	1379	1281	8.4	65.7	82	402	117	157	873.0
kegg_metabolic_cje	851	785	8.7	55.8	70	269	77	101	544.6
kegg_metabolic_cmu	650	608	10.3	55.1	57	226	63	79	431.6
kegg_metabolic_cpa	659	618	10.1	54.9	55	226	64	81	436.4
kegg_metabolic_cpe	1068	979	8.9	62.7	85	335	89	116	682.2

kegg_metabolic_cpj	660	619	10.1	55.0	55	227	67	82	437.8
kegg_metabolic_cpn	660	619	10.1	55.0	55	226	64	80	437.2
kegg_metabolic_cte	1000	908	8.3	57.5	75	311	72	92	630.8
kegg_metabolic_ctr	652	610	10.3	55.4	57	227	63	79	433.5
kegg_metabolic_dme	1372	1262	9.0	68.7	81	408	116	148	865.7
kegg_metabolic_dra	1185	1101	8.9	65.0	85	367	97	131	756.7
kegg_metabolic_ecc	1533	1407	8.8	71.3	96	444	122	166	956.4
kegg_metabolic_ece	1604	1469	8.6	71.1	94	458	119	164	994.9
kegg_metabolic_ecj	1635	1489	8.7	72.2	96	468	116	160	1009.6
kegg_metabolic_eco	1641	1495	8.7	72.1	97	469	116	159	1013.0
kegg_metabolic_ecs	1601	1463	8.6	71.0	94	457	117	163	991.4
kegg_metabolic_fnu	1031	946	8.9	62.0	79	332	87	115	664.3
kegg_metabolic_hal	808	744	8.5	53.1	61	256	76	93	516.9
kegg_metabolic_hin	1085	1014	8.8	62.7	90	337	99	128	697.0
kegg_metabolic_hpj	849	772	8.9	56.8	71	278	74	96	547.1
kegg_metabolic_hpy	863	780	8.9	57.1	74	282	74	96	554.0
kegg_metabolic_hsa	1917	1732	8.6	75.4	95	502	167	215	1157.2
kegg_metabolic_lil	1234	1130	9.9	70.4	85	387	110	138	786.4
kegg_metabolic_lin	980	918	8.7	59.2	82	300	92	119	625.9
kegg_metabolic_lla	966	892	9.3	61.5	84	305	94	118	619.5
kegg_metabolic_lmo	1004	937	8.7	59.8	87	308	90	121	640.4
kegg_metabolic_mac	938	845	8.6	57.1	66	296	78	106	594.7
kegg_metabolic_mge	358	327	9.2	39.4	43	131	31	37	236.0
kegg_metabolic_mja	796	736	8.6	53.9	65	261	71	89	517.3
kegg_metabolic_mka	754	690	8.1	50.7	72	250	65	79	488.4
kegg_metabolic_mle	1008	944	8.8	60.5	78	323	83	104	651.2
kegg_metabolic_mlo	1654	1499	8.5	71.4	92	470	121	167	1017.6
kegg_metabolic_mma	1008	911	8.4	58.3	67	318	81	109	640.7
kegg_metabolic_mmu	1863	1682	8.5	74.4	87	487	169	216	1125.4
kegg_metabolic_mpe	579	537	9.3	48.9	47	198	58	76	380.6
kegg_metabolic_mpn	375	347	9.5	41.1	45	136	34	42	248.7
kegg_metabolic_mpu	385	353	8.9	40.2	37	137	33	45	253.3
kegg_metabolic_mtc	1443	1340	8.9	69.4	91	430	108	144	909.8
kegg_metabolic_mth	814	749	8.0	52.1	58	263	72	92	527.5
kegg_metabolic_mtu	1509	1397	8.7	70.2	92	450	108	148	950.4
kegg_metabolic_nma	1017	944	8.7	59.7	77	313	94	121	649.9
kegg_metabolic_nme	1020	949	8.7	60.2	80	316	91	118	653.7
kegg_metabolic_oih	1340	1237	9.2	69.2	91	409	102	137	849.9
kegg_metabolic_pab	829	748	8.8	55.2	63	277	64	81	529.6
kegg_metabolic_pae	1581	1450	8.4	70.0	101	454	123	161	984.8
kegg_metabolic_pai	958	881	8.9	59.1	64	316	73	91	616.9
kegg_metabolic_pfu	814	735	8.6	53.9	72	265	69	81	517.1
kegg_metabolic_pho	733	662	8.5	51.0	63	248	57	69	470.8
kegg_metabolic_pmu	1167	1091	8.6	62.6	93	348	107	141	740.2
kegg_metabolic_rco	785	710	9.4	55.3	68	266	66	81	504.3
kegg_metabolic_rno	1646	1483	8.3	69.9	82	449	140	183	1006.6
kegg_metabolic_rpr	754	688	9.3	54.2	69	257	63	78	487.6
kegg_metabolic_rso	1642	1505	8.6	71.7	91	473	126	168	1024.6
kegg_metabolic_sag	875	808	8.3	55.4	70	278	90	111	564.5
kegg_metabolic_sam	1073	1012	8.8	61.8	80	323	110	141	685.2
kegg_metabolic_san	880	812	8.3	55.6	70	280	91	112	568.1
kegg_metabolic_sau	1080	1016	8.9	61.9	81	325	110	141	689.1
kegg_metabolic_sav	1072	1010	8.8	61.6	81	322	110	141	683.9
kegg_metabolic_sce	1378	1265	8.5	65.8	75	387	117	158	853.1
kegg_metabolic_sco	1542	1413	8.7	69.9	92	447	107	145	956.0
kegg_metabolic_sfl	1560	1439	8.9	72.0	97	455	119	165	977.4

kegg_metabolic_sme	1705	1545	8.3	71.8	94	493	114	159	1053.1
kegg_metabolic_smu	911	835	9.3	59.6	82	288	90	113	582.3
kegg_metabolic_son	1417	1306	8.7	68.3	87	417	117	151	888.7
kegg_metabolic_spg	822	761	9.0	56.6	68	271	83	106	534.3
kegg_metabolic_spm	820	759	9.0	56.4	70	269	82	107	532.6
kegg_metabolic_spn	892	829	8.6	57.1	67	286	94	117	578.5
kegg_metabolic_spo	1184	1115	8.4	61.6	71	341	114	144	748.0
kegg_metabolic_spr	910	848	8.6	57.6	67	291	91	116	589.6
kegg_metabolic_spy	825	763	9.0	56.7	67	272	81	107	535.9
kegg_metabolic_sso	1035	949	8.6	60.2	88	339	81	98	665.5
kegg_metabolic_stm	1563	1439	8.8	72.1	98	454	122	169	975.9
kegg_metabolic_sto	973	900	8.7	59.4	84	322	81	103	632.5
kegg_metabolic_sty	1547	1420	8.8	71.8	91	450	123	167	965.7
kegg_metabolic_syn	1264	1157	8.8	65.3	88	379	110	142	797.5
kegg_metabolic_tac	768	701	8.5	52.1	76	250	62	76	488.9
kegg_metabolic_tel	1159	1056	9.0	64.1	81	358	92	123	733.7
kegg_metabolic_tma	914	846	9.4	59.7	70	287	83	108	584.3
kegg_metabolic_tpa	524	487	7.7	42.7	41	181	48	63	345.4
kegg_metabolic_tte	991	914	8.6	59.5	72	310	89	115	634.9
kegg_metabolic_tvo	778	714	8.4	51.7	64	252	64	78	496.2
kegg_metabolic_uur	321	285	8.1	35.0	31	121	21	27	208.5
kegg_metabolic_vch	1422	1313	8.7	68.8	95	419	114	155	894.7
kegg_metabolic_wbr	775	725	8.5	51.9	67	240	75	94	495.2
kegg_metabolic_xac	1444	1312	8.8	69.6	90	433	106	143	902.5
kegg_metabolic_xcc	1454	1323	8.8	69.8	89	433	107	145	908.7
kegg_metabolic_xfa	1092	1002	9.1	62.9	78	336	90	115	689.8
kegg_metabolic_ype	1474	1355	8.7	69.7	100	431	118	162	922.5
kegg_metabolic_ypk	1452	1338	8.7	69.3	94	429	119	159	913.9
kidnappings	351	94	6.3	19.9	6	48	5	6	76.9
law_firm	71	71	2.5	8.6	5	23	4	7	45.7
lesmis	77	64	2.8	7.1	12	10	12	15	28.4
london_transport	369	358	35.0	92.8	67	184	11	22	287.6
macaque_neural	47	47	2.3	6.8	3	14	5	7	29.4
macaques	62	59	2.6	8.1	7	18	7	8	35.2
malaria_genes_HVR_1	307	307	4.3	22.4	26	79	45	58	190.5
malaria_genes_HVR_2	307	290	5.0	22.8	29	56	42	48	182.4
malaria_genes_HVR_3	307	255	4.7	23.4	29	118	27	28	186.5
malaria_genes_HVR_4	307	282	7.0	28.2	35	69	48	56	186.2
malaria_genes_HVR_5	307	298	7.3	32.0	33	95	39	52	197.2
malaria_genes_HVR_6	307	304	4.0	20.9	21	77	35	47	185.0
malaria_genes_HVR_7	307	306	2.4	10.7	17	19	69	79	122.8
malaria_genes_HVR_8	307	295	2.6	14.0	12	46	49	63	160.7
malaria_genes_HVR_9	307	305	2.4	13.4	17	47	55	68	158.6
malaria_genes_combined	307	307	1.9	12.6	14	63	18	25	182.0
marvel_partnerships	350	283	30.7	81.9	45	179	15	24	247.6
marvel_universe	19428	9376	40.0	374.0	522	2268	1254	1560	5892.4
messal_shale	700	421	8.5	42.5	32	149	46	63	287.7
montreal	35	23	2.4	5.8	3	12	1	1	17.8
moreno_sheep	28	27	2.0	4.7	2	8	4	5	15.4
mouse_control_rnn	669	669	20.1	84.8	60	248	66	102	457.4
mouse_meso	213	185	3.6	11.8	27	20	64	70	64.4
mouse_rnn	178	178	1.2	4.9	6	7	41	48	63.2
mouse_voxel	15314	15313	8.0	62.1	416	116	2221	2741	2169.1
netscience	1589	1416	21.5	120.7	122	532	136	196	973.5
new_guinea_tribes	16	16	2.0	4.2	1	7	1	1	11.4
new_zealand_collab	1511	678	1.6	3.8	21	4	139	153	25.9

non_financial_institution04-Jan-2001	155	155	2.6	12.2	22	52	12	14	102.1
november17	22	20	2.1	4.9	3	9	1	2	14.9
openflights	3214	2139	5.8	48.5	117	209	467	529	915.4
packet_delays	10	10	1.0	1.2	1	1	1	1	2.2
pdumerilii_desmosomal	2807	2415	76.1	281.7	424	644	647	756	1419.6
pdumerilii_neuronal	2544	1592	14.3	78.9	177	247	788	870	796.3
physician_trust	241	211	19.1	50.4	23	95	9	17	159.7
physics_collab_arXiv	14488	13536	38.8	390.4	722	2296	2263	2690	7339.7
physics_collab_pierreAuger	514	510	10.0	39.4	59	93	110	124	251.9
plant_pol_kato	772	150	3.7	11.5	14	20	36	40	64.2
plant_pol_robertson	1884	822	9.8	60.0	44	265	96	134	540.1
plant_pol_vazquez_All_sites_pooled	144	29	2.9	4.2	4	4	10	10	8.7
plant_pol_vazquez_Arroyo_Goye	144	20	2.3	3.6	2	3	8	8	8.2
plant_pol_vazquez_Cerro_Lopez	144	18	3.2	5.0	4	4	4	4	10.6
plant_pol_vazquez_Llao_Llao	144	20	3.0	4.6	8	6	8	8	8.7
plant_pol_vazquez_Mascardi_c	144	16	3.6	5.3	8	6	4	4	9.3
plant_pol_vazquez_Mascardi_nc	144	16	2.4	3.5	4	4	6	6	7.2
plant_pol_vazquez_Quetrihue_c	144	16	2.6	4.4	2	5	2	2	9.6
plant_pol_vazquez_Quetrihue_nc	144	14	2.3	2.8	4	3	4	4	4.3
plant_pol_vazquez_Safariland	144	18	2.1	2.6	4	2	4	4	4.3
polblogs	1490	784	6.0	41.7	28	193	167	199	454.9
polbooks	105	105	6.2	18.8	12	41	9	13	74.3
power	4941	4348	235.5	860.9	351	2278	49	148	3584.4
qa_user_mathoverflow_c2a	13840	5559	3.3	32.9	186	117	1421	1562	1005.9
qa_user_mathoverflow_c2q	16836	11155	6.7	86.9	352	356	2221	2558	3646.8
reactome	6327	4654	6.9	63.6	162	176	716	840	1922.9
reality_mining	96	96	3.0	9.2	16	16	22	26	42.1
residence_hall	217	215	9.9	34.3	24	85	16	24	152.5
rhesus_monkey	16	16	1.7	3.4	4	5	2	3	9.4
sp_high_school_diaries	329	115	12.6	30.8	20	55	6	10	88.8
sp_high_school_facebook	329	156	4.8	17.2	9	44	16	24	96.2
sp_high_school_new_2011	126	126	11.0	22.3	33	28	31	36	62.2
sp_high_school_new_2012	180	180	6.0	19.4	20	40	44	50	90.4
sp_high_school_proximity	329	327	19.6	51.5	67	89	67	81	182.8
sp_high_school_survey	329	130	8.6	25.6	16	57	8	13	97.1
sp_hospital	75	75	2.7	7.2	12	11	21	24	29.5
sp_hypertext_contacts	113	113	3.9	11.0	19	19	26	30	49.7
sp_hypertext_intervals	113	113	2.0	9.8	2	42	2	5	77.1
sp_infectious	10972	10922	202.8	1011.9	978	3437	1531	1948	6672.5
sp_kenyan_households	47	47	2.6	6.6	10	11	10	12	23.9
sp_office	92	92	4.0	10.7	14	19	26	28	43.0
sp_primary_school_day_1	236	236	9.1	32.8	28	86	18	27	158.4
sp_primary_school_day_2	238	238	8.5	30.5	26	77	26	38	150.8
student_cooperation	185	184	23.9	54.4	34	82	16	28	144.8
swingers	96	38	2.8	7.7	3	16	3	3	27.9
terrorists_911	62	54	4.0	11.3	9	23	5	7	39.9
train_terrorists	64	63	2.0	6.8	7	16	9	12	37.2
ugandan_village_friendship-1	203	200	15.3	43.1	33	84	16	23	145.3
ugandan_village_friendship-10	207	207	9.3	33.7	20	86	13	23	149.9
ugandan_village_friendship-11	250	250	13.9	44.3	25	98	23	35	174.2
ugandan_village_friendship-12	229	229	8.5	32.2	23	86	23	34	156.2
ugandan_village_friendship-13	183	183	10.0	32.5	29	74	16	24	128.5
ugandan_village_friendship-14	124	124	7.8	23.8	16	51	10	16	88.9
ugandan_village_friendship-15	120	116	10.6	27.2	15	49	8	13	83.5
ugandan_village_friendship-16	372	372	19.4	64.9	36	150	31	49	264.3
ugandan_village_friendship-17	65	65	3.7	11.2	7	25	6	7	44.4

ugandan_village_friendship-2	182	182	8.9	30.0	27	71	15	21	126.2
ugandan_village_friendship-3	192	192	7.5	27.9	13	72	21	30	131.2
ugandan_village_friendship-4	320	320	12.4	47.3	32	126	28	43	224.7
ugandan_village_friendship-5	184	184	10.2	32.7	22	72	17	26	129.5
ugandan_village_friendship-6	139	139	7.3	24.1	12	56	10	19	98.1
ugandan_village_friendship-7	121	121	6.6	21.7	12	51	5	11	88.0
ugandan_village_friendship-8	369	369	16.6	60.3	48	153	25	45	267.1
ugandan_village_friendship-9	178	178	6.4	24.6	16	65	18	27	122.2
ugandan_village_health-advice_1	190	161	7.8	26.9	19	67	11	19	117.4
ugandan_village_health-advice_10	205	202	6.8	26.7	18	74	20	31	135.7
ugandan_village_health-advice_11	240	187	7.8	28.6	22	74	18	28	134.7
ugandan_village_health-advice_12	221	210	6.1	26.4	19	85	15	22	149.7
ugandan_village_health-advice_13	173	155	10.4	32.3	18	75	4	8	119.5
ugandan_village_health-advice_14	120	118	9.3	25.1	18	47	12	15	83.3
ugandan_village_health-advice_15	117	77	6.3	17.3	12	33	6	10	59.0
ugandan_village_health-advice_16	350	310	5.6	28.8	16	114	22	32	211.7
ugandan_village_health-advice_17	63	58	5.1	13.1	11	24	5	8	41.8
ugandan_village_health-advice_2	170	134	6.0	20.3	18	50	11	16	92.4
ugandan_village_health-advice_3	185	139	5.9	20.3	20	47	19	24	93.6
ugandan_village_health-advice_4	316	310	13.7	49.2	26	123	25	43	218.8
ugandan_village_health-advice_5	168	149	6.1	22.0	17	56	15	22	104.1
ugandan_village_health-advice_6	134	95	4.1	14.2	13	38	7	9	68.1
ugandan_village_health-advice_7	121	121	6.1	20.2	17	48	10	15	85.2
ugandan_village_health-advice_8	361	358	15.5	55.8	31	141	28	46	249.7
ugandan_village_health-advice_9	173	155	5.2	19.7	20	54	18	23	101.4
un_migrations	232	232	1.9	5.8	19	10	93	96	36.4
uni_email	1133	1091	25.3	117.1	122	365	152	201	708.6
unicodelang	868	374	7.1	33.6	24	141	14	22	233.0
us_agencies_alabama	1281	536	1.8	4.3	19	3	202	212	30.2
us_agencies_alaska	557	264	1.6	2.7	15	2	98	102	8.8
us_agencies_arizona	382	218	1.9	3.7	9	4	78	82	12.6
us_agencies_arkansas	561	273	1.6	3.4	14	4	105	110	12.5
us_agencies_california	4292	2254	5.8	17.3	87	19	858	901	122.3
us_agencies_colorado	1095	588	1.0	1.7	9	1	202	209	7.0
us_agencies_connecticut	685	406	1.6	4.7	25	7	151	158	31.8
us_agencies_delaware	446	230	1.6	3.8	9	3	86	88	18.2
us_agencies_florida	2263	890	1.3	3.4	27	2	320	337	27.6
us_agencies_georgia	1092	652	1.5	2.6	14	2	233	243	9.3
us_agencies_hawaii	234	71	1.1	1.6	6	1	24	26	4.4
us_agencies_idaho	1109	572	1.2	3.5	28	4	239	245	31.1
us_agencies_illinois	1133	678	1.1	2.2	12	1	225	238	12.7
us_agencies_indiana	1385	620	1.3	3.6	15	3	251	258	33.9
us_agencies_iowa	541	300	4.5	8.5	24	8	105	112	25.9
us_agencies_kansas	675	355	1.7	3.4	11	3	129	136	14.6
us_agencies_kentucky	1049	520	1.0	1.6	6	1	200	208	6.1
us_agencies_louisiana	715	272	3.2	5.2	16	4	97	102	15.5
us_agencies_maine	547	330	4.1	9.2	30	10	130	131	35.8
us_agencies_maryland	564	325	2.0	4.1	14	4	112	120	15.7
us_agencies_massachusetts	1466	1021	3.1	8.3	48	8	413	419	53.0
us_agencies_michigan	1796	981	2.4	4.8	20	3	330	349	25.8
us_agencies_minnesota	1527	877	1.0	1.3	8	1	309	322	4.7
us_agencies_mississippi	408	185	1.8	2.8	9	2	64	67	8.6
us_agencies_missouri	1161	549	1.8	4.1	22	4	187	197	20.3
us_agencies_montana	243	108	1.0	1.6	6	1	42	43	4.3
us_agencies_nebraska	550	285	1.5	3.2	8	3	103	105	13.9
us_agencies_nevada	388	227	2.6	5.5	20	6	91	94	21.8

us_agencies_newhampshire	675	404	1.0	1.4	6	1	144	149	3.5
us_agencies_newjersey	1606	942	1.8	3.3	18	2	356	369	17.2
us_agencies_newmexico	411	255	1.2	2.2	8	2	94	99	8.6
us_agencies_newyork	2452	1454	2.1	5.2	30	5	486	510	34.7
us_agencies_northcarolina	974	617	1.0	2.0	5	1	211	222	14.0
us_agencies_northdakota	515	292	1.1	2.0	7	1	108	111	9.9
us_agencies_ohio	1153	680	1.4	3.2	23	3	214	228	18.6
us_agencies_oklahoma	736	289	1.6	3.4	13	4	108	112	12.8
us_agencies_oregon	522	312	2.3	5.3	13	5	117	123	24.2
us_agencies_pennsylvania	1317	554	2.3	3.1	12	3	186	197	6.5
us_agencies_rhodeisland	294	186	1.8	4.0	12	5	68	70	15.2
us_agencies_southcarolina	1530	730	1.4	3.0	12	4	307	311	15.7
us_agencies_southdakota	403	193	1.3	2.8	10	3	70	71	12.2
us_agencies_tennessee	363	189	1.0	1.3	4	1	67	69	3.0
us_agencies_texas	1079	689	1.1	1.9	10	1	255	265	9.4
us_agencies_utah	1158	431	1.2	3.0	22	3	171	179	20.5
us_agencies_vermont	469	237	1.1	1.9	11	2	97	99	6.9
us_agencies_virginia	560	391	2.3	6.0	35	7	139	147	29.4
us_agencies_washington	999	617	1.0	1.2	6	1	227	237	3.1
us_agencies_westvirginia	579	247	1.1	1.9	8	1	79	86	9.7
us_agencies_wisconsin	1371	904	1.6	5.2	22	7	323	335	46.3
us_agencies_wyoming	318	199	1.0	1.1	2	1	89	90	2.0
us_air_traffic	2278	1758	1.5	9.2	57	21	734	755	195.0
us_congress_H100	446	446	1.7	13.0	15	82	21	35	257.1
us_congress_H101	449	449	1.6	12.1	14	70	23	37	251.6
us_congress_H102	447	447	1.5	11.2	16	56	23	37	240.5
us_congress_H103	446	446	1.5	10.8	15	48	25	37	236.5
us_congress_H104	445	445	1.4	9.9	15	42	26	42	231.7
us_congress_H105	449	449	1.4	9.3	7	33	30	47	227.1
us_congress_H106	442	442	1.4	9.5	12	37	27	46	227.0
us_congress_H107	447	447	1.4	9.8	13	39	26	43	230.5
us_congress_H108	444	444	1.3	8.9	14	27	32	49	217.8
us_congress_H109	445	445	1.3	8.9	14	26	32	51	217.7
us_congress_H110	452	452	1.3	8.9	10	26	34	52	222.6
us_congress_H111	451	451	1.3	9.0	11	28	29	46	222.8
us_congress_H112	450	450	1.2	7.2	12	6	42	63	192.0
us_congress_H113	447	447	1.2	7.5	10	8	38	58	196.9
us_congress_H114	446	446	1.2	7.3	7	7	44	64	192.0
us_congress_H93	446	446	3.0	22.3	13	135	41	64	281.2
us_congress_H94	445	445	2.8	21.6	14	137	36	57	284.6
us_congress_H95	444	444	2.8	22.7	12	151	23	46	297.1
us_congress_H96	442	442	1.6	12.5	13	85	20	32	260.2
us_congress_H97	447	447	1.7	12.7	14	80	20	31	257.5
us_congress_H98	444	444	1.7	13.0	16	81	20	34	255.7
us_congress_H99	443	443	1.7	12.7	12	76	20	33	251.2
us_congress_S100	101	101	1.9	8.3	8	29	7	8	62.5
us_congress_S101	101	101	1.8	7.5	5	25	7	10	60.3
us_congress_S102	102	102	1.8	7.6	6	25	7	11	60.5
us_congress_S103	101	101	1.8	7.8	5	27	6	9	61.2
us_congress_S104	102	102	1.7	7.4	5	26	5	10	61.1
us_congress_S105	100	100	1.5	6.0	5	19	6	9	54.6
us_congress_S106	102	102	1.6	6.7	4	23	6	9	59.2
us_congress_S107	101	101	1.7	7.4	5	26	5	8	60.3
us_congress_S108	100	100	1.8	7.8	4	28	4	8	61.7
us_congress_S109	101	101	1.6	6.7	6	22	6	9	57.6
us_congress_S110	102	102	1.6	6.7	6	23	6	9	58.4

us_congress_S111	109	109	1.4	5.9	6	18	6	9	57.5
us_congress_S112	101	101	1.2	4.3	4	6	9	15	44.6
us_congress_S113	105	105	1.3	4.7	6	7	9	15	48.5
us_congress_S114	100	100	1.3	4.5	6	6	7	11	45.5
us_congress_S93	101	101	1.7	7.5	6	29	5	6	62.9
us_congress_S94	100	100	1.7	7.3	6	27	6	8	61.0
us_congress_S95	104	104	1.8	8.2	6	32	5	6	66.5
us_congress_S96	101	101	1.7	7.9	4	30	4	7	63.9
us_congress_S97	101	101	2.0	8.5	8	31	6	7	63.6
us_congress_S98	101	101	1.9	8.4	6	30	5	8	63.6
us_congress_S99	101	101	1.9	8.6	4	33	4	7	64.9
webkb_webkb_cornell_cocite	346	344	2.2	10.8	17	23	72	84	135.9
webkb_webkb_cornell_link1	349	253	9.4	34.4	24	87	28	42	171.9
webkb_webkb_texas_cocite	334	334	2.6	11.5	21	20	66	76	120.4
webkb_webkb_texas_link1	286	220	12.2	37.5	39	82	21	28	155.3
webkb_webkb_washington_cocite	434	434	2.5	10.5	24	14	95	109	124.8
webkb_webkb_washington_link1	433	321	7.7	32.4	32	96	32	44	206.2
webkb_webkb_wisconsin_cocite	348	348	1.7	8.5	15	19	78	91	118.7
webkb_webkb_wisconsin_link1	300	232	8.9	33.2	30	85	28	36	164.7
wiki_rfa	11381	3475	10.9	123.3	128	993	456	600	2235.7
wiki_science	687	641	7.1	40.0	42	159	82	106	383.1
wiki_talk_br	1181	428	1.8	5.0	23	4	108	116	43.9
wiki_talk_cy	2233	610	1.5	5.7	25	4	140	151	105.3
wiki_talk_eo	7586	1553	3.2	13.2	55	13	345	433	211.0
wiki_talk_gl	8097	1061	2.0	8.0	40	9	276	292	130.6
wiki_talk_ht	536	197	1.2	4.5	9	3	25	25	69.9
wiki_talk_oc	3144	452	1.8	5.6	21	5	81	88	62.8
wikipedia_link_bcl	8130	5815	2.4	32.2	146	64	634	775	2240.6
wikipedia_link_bh	15581	8008	2.6	40.6	140	70	666	829	3400.5
wikipedia_link_bug	14266	13579	9.8	33.6	53	23	53	66	1287.6
wikipedia_link_cdo	14816	9581	4.3	83.5	183	637	1135	1387	4670.0
wikipedia_link_co	8252	4648	3.9	44.9	97	92	471	602	2256.1
wikipedia_link_crh	8286	4923	9.0	64.4	145	46	346	425	2049.2
wikipedia_link_csb	5561	2509	1.2	13.4	30	5	240	335	1236.6
wikipedia_link_diq	11810	7380	5.0	65.9	176	174	882	1102	3326.7
wikipedia_link_dv	4266	2242	1.5	11.5	24	2	229	297	887.3
wikipedia_link_eml	11856	8712	1.2	15.8	89	15	1377	1641	2959.8
wikipedia_link_fiu_vro	6424	3795	1.9	21.8	56	18	424	538	1691.8
wikipedia_link_fo	17649	12042	4.8	116.7	280	1728	1707	2100	6675.2
wikipedia_link_gag	2929	1721	2.6	20.7	57	31	272	316	622.6
wikipedia_link_gan	9189	5971	9.9	106.8	213	456	617	827	3015.2
wikipedia_link_glk	7332	5457	4.0	41.6	239	104	383	442	1873.7
wikipedia_link_gv	6862	4814	10.0	93.7	181	380	623	777	2372.0
wikipedia_link_hak	11487	7548	6.1	58.1	163	49	765	985	3107.2
wikipedia_link_hif	8758	3667	2.4	24.0	66	22	595	731	1439.8
wikipedia_link_kv	6914	4878	3.3	45.5	98	236	896	1055	2280.8
wikipedia_link_lez	5171	3753	7.4	54.1	102	68	602	712	1480.0
wikipedia_link_lo	3811	1937	1.1	11.5	24	4	185	241	956.4
wikipedia_link_map_bms	14162	6667	4.3	54.8	204	181	943	1077	2641.0
wikipedia_link_mhr	14569	7921	7.2	110.2	219	876	950	1195	4417.3
wikipedia_link_myv	6293	4533	9.2	82.8	134	280	662	797	2152.0
wikipedia_link_mzn	18112	10510	5.2	40.8	169	26	770	934	2986.4
wikipedia_link_nah	10285	6646	13.7	138.7	305	670	798	983	3491.3
wikipedia_link_nap	15441	12924	9.9	91.9	224	77	1737	2245	4489.6
wikipedia_link_nds_nl	10453	7592	22.3	226.5	362	1458	1043	1297	4369.4
wikipedia_link_nso	8152	7497	3.4	45.0	186	92	855	1253	2911.0

wikipedia_link_pam	9537	6122	8.8	88.3	252	225	878	1041	2685.5
wikipedia_link_ps	12949	6086	1.8	21.5	83	14	662	799	2388.0
wikipedia_link_roa_tara	10423	9415	10.6	67.4	194	60	861	1173	2399.3
wikipedia_link_rue	7621	5552	2.7	48.3	85	419	631	777	2919.7
wikipedia_link_sah	15531	10209	2.5	58.6	139	611	1194	1511	5548.8
wikipedia_link_se	8622	5846	12.9	89.3	192	97	732	913	2391.4
wikipedia_link_so	7439	4580	3.8	56.7	136	408	544	680	2399.3
wikipedia_link_tk	6138	4724	2.6	40.5	97	223	533	620	2280.7
wikipedia_link_vls	9941	7171	8.0	133.0	185	1586	888	1154	4325.3
wikipedia_link_xal	2697	1666	1.7	8.8	16	4	84	110	451.6
wikipedia_link_xmf	15600	10214	4.8	89.7	250	826	1544	1876	5091.0
windsurfers	43	43	1.7	4.3	5	6	8	10	18.9
word_adjacency_darwin	7381	2698	10.3	94.7	181	486	402	484	1594.2
word_adjacency_french	8325	2717	16.9	130.1	197	684	304	386	1819.6
word_adjacency_spanish	11586	2794	10.7	92.9	170	423	415	499	1634.0
yeast_transcription	916	123	10.9	29.6	18	58	8	10	97.8
zebrafish_meso	71	71	1.8	6.5	6	17	6	10	41.5
zebrafish_rnn	4589	4589	16.5	160.3	262	943	1103	1258	2336.3
zebras	27	26	1.5	4.0	1	8	4	5	17.4