The low-rank hypothesis of complex systems

Vincent Thibeault,1,2,∗ Antoine Allard,1,2,† and Patrick Desrosiers1,2,3,‡

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

Complex systems are high-dimensional nonlinear dynamical systems with intricate interactions among their constituents. To make interpretable predictions about their large-scale behavior, it is typically assumed, without a clear statement, that these dynamics can be reduced to a few number of equations involving a low-rank matrix describing the network of interactions—what we call the low-rank hypothesis. By leveraging fundamental theorems on singular value decomposition, we verify the hypothesis for various random networks, either by making explicit their low-rank formulation or by demonstrating the exponential decrease of their singular values. Notably, we validate the hypothesis experimentally for real networks by showing that their effective rank is considerably lower than their number of vertices. 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 to connect the rapidly decreasing singular values of real networks to the dimension reduction error of the nonlinear dynamics they support, be it microbial, neuronal or epidemiological. Finally, we prove that higher-order interactions naturally emerge from the dimension reduction, thus providing theoretical insights into the origin of higher-order interactions in complex systems.

Unraveling the emergent phenomena that drive the functions of complex systems requires to rally the microscopic mechanisms with the macroscopic ones. Rather than decomposing complex systems in 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 an appropriate dimension for a low-dimensional model while preserving the essential features of the original system 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 behavior”: the logistic equation [6], cellular automata [7, 8], or spin glasses [9, 10] exhibit complex behaviors such as chaos, and highly idealized recurrent neural networks 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 simplified to a discrete mathematical structure—typically a graph, defined by a set of vertices and edges (see Figs. 1a and 1b). Such representation allows for the extraction of some dominant properties of complex systems, such as their organization into modules [13]. A graph can always be described as a matrix or a tensor. This simple, yet important, possibility unlocks several tools and concepts from linear algebra to characterize networks. Among them, spectral theory allows identifying the fundamental components of matrices through matrix decomposition. Eigenvalue decomposition has long been used to extract key properties of graphs, such as their invariants [14], their modular structure [15], the centrality of their vertices [16], or the bifurcations of dynamical systems taking place on these networks [17].

One pressing challenge nowadays in network science is to efficiently adapt the tools of spectral theory to directed, weighted, and signed (e.g., excitatory-inhibitory) networks and hence, to general real matrices. Indeed, a direct use of the standard eigenvalue decomposition yields complex eigenvalues and complex-valued eigenvectors, which are generally hard to relate to intuitive properties of networks. For example, it is unclear how to use them to define a vertex centrality measure (see SI II D) or observables for dimension reductions (see SI II F). Even worse, it is not even guaranteed that the matrix representation of the network is diagonalizable. For instance, the trivial directed graph with two vertices connected by one directed edge is not diagonalizable. Also, any network whose (real) matrix representation, \( W \), is rectangular is not diagonalizable (e.g., incidence matrix, interlayer adjacency matrix in multilayer networks).

Yet, the matrices \( WW^\top \) and \( W^\top W \) are always square and symmetric. By the spectral theorem, they are thus both diagonalizable, which lays the foundations of singular value decomposition (SVD, see Theorem S4), as illustrated in Fig. 1c. Interestingly, the decomposition exists for any matrix, the singular vectors are real-valued, and the singular values \( \sigma_1, ..., \sigma_N \) are nonnegative real
numbers. Notably, the number of nonzero singular values equals the rank of $W$, also commonly defined as the maximal number of linearly independent rows or columns of the matrix. Moreover, because eigenvalue decomposition is used to construct it, SVD inherits many similar theorems to the eigenvalue decomposition [18], such as Weyl’s theorem (see Theorem S8), but it also opens the door to new fundamental results. In particular, SVD is a central tool for dimension reduction in general: the Schmidt-Eckart-Young-Mirsky theorem guarantees that the truncated SVD yields the best low-rank approximation of a matrix (see Fig. 1c and Theorem S10).

The salient properties of SVD and its close relationship with the rank of a matrix have not yet been completely recognized in network science and spectral graph theory if we compare to its ubiquity in data science (e.g., matrix completion [19], dynamic mode decomposition [20], and optimal singular value shrinkage [21]), control theory (e.g., Kalman criterion [22, 23]), random matrix theory (e.g., Marčenko-Pastur’s law [24]), and linear algebra (e.g., matrix norms [18]). SVD is not even mentioned in many of the main introductory textbooks of network science [25] or spectral graph theory [26].

Throughout the paper, we leverage the key attributes of SVD to define and evaluate the impact of the low-rank hypothesis of complex systems. Before tackling the case of complex systems as high-dimensional nonlinear dynamical systems, we first report theoretical and empirical evidence supporting the hypothesis for networks.

**Evidence of the low-rank hypothesis for networks**

As a first step, it is especially instructive to consider random networks. A random network is a set of networks equipped with a probability measure that depends on some properties, such as the degrees, the modules, or the distance between vertices in a hyperbolic space (see SI II B and SI II C). Mathematically, it can always be written as a random matrix $W = \langle W \rangle + X$, where $\langle W \rangle$ is the expected weight matrix and $X$ is a random matrix with mean 0.

Through the lens of SVD, we found that the expected matrices from many widely used random graphs involve low-rank matrices. Indeed, we highlight the—usually implicit—assumption that $\langle W \rangle$ is a function of a low-rank matrix (see Table I in Methods). In many cases, it is straightforward to see the low rank of $\langle W \rangle$ since it can be written into its rank factorized form (see SI II B), which is already a first simple indicator of the low-rank
hypothesis for networks.

However, the low rank of \( (W) \) is not always obvious, such as in the case of the directed soft configuration model. Using Weyl’s inequalities for singular values [27, 28], we demonstrated in Theorem 1 (Methods) that the singular values of \( (W) \) for the directed soft configuration model are bounded above by an exponentially decreasing term. This decreasing tendency is also observed in many random networks for the instances of \( W \), which are generally of full rank (i.e., all singular values are nonzero) (see Fig. S3). The rapid decrease of the singular values hints at the approximate low rank of a network, which is a second crucial indicator of the low-rank hypothesis for networks.

The attributes “rapid decrease” and “approximately of low rank” remain to be quantified, however. To do so, we invoke the notion of effective ranks, and first consider the stable rank: a measure of the relative importance of the squared singular values with respect to the squared largest singular value (see Table II in Methods). It is stable in the sense that it remains essentially unchanged under small perturbations of the matrix, contrarily to the rank. The stable rank tends to be low when a matrix is approximately of low rank: as described in SI II A, we demonstrated that having exponentially decreasing singular values (such as in the directed soft configuration model) is a sufficient criterion for the stable rank to be of order \( O(1) \) as the number of vertices \( N \) tends to infinity. Having low effective ranks compared to \( N \) is thus a third indicator, this time quantitative, of the low-rank hypothesis for networks.

Despite its frequent use—often implicit, but sometimes very explicit [29, 30]—the low-rank hypothesis has yet to be verified experimentally for real networks in all their diversity. Our experiments indeed 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 \( \text{Drosophila melanogaster} \) in Fig. 1d. Figure 1e presents a broader overview of the low-rank hypothesis for networks by gathering 674 real networks from 10 different origins, from connectomes and interactomes to ecological and social networks. Clearly, the stable rank of most networks is considerably lower than their rank.

To ensure that this observation is not specific to the stable rank, we report in Figs. 1f to 1i that the same conclusion holds despite the specifics of the definition of the effective rank although some definitions yield larger values. We also report that the effective ranks all tend to grow linearly, with a small slope, with respect to the size of the networks (see Fig. S2). Contrarily to the effective ranks, the rank of networks is often comparable to their dimension (see Fig. 1m). This observation is expected, especially for weighted networks with real elements, since non-invertible matrices form a set of measure 0. On the contrary, undirected binary networks are more likely to have linearly dependent rows/columns, or, in other words, vertices with the same neighborhood.

The low-rank hypothesis is therefore justifiable for real networks: we observe rapid decreases of their singular values and their effective ranks are considerably lower than their size. Interestingly, these observations seem to be ubiquitous for big data matrices [31–33], but it remains a puzzling phenomenon. In particular, the consequences of these observations for high-dimensional nonlinear dynamics on networks are still to be untangle, which is the subject of the next section.

**The induced low-dimensional hypothesis**

Intuitively, we expect the low (effective) rank of networks to guide the dimension reduction of dynamics on these networks and to help establish an appropriate dimension at which emergent phenomena take place in complex systems. To evaluate its role, let us first consider a general class of \( N \)-dimensional (complete) dynamics on network described by the differential equations \( \dot{x} = f(x; W) \), where \( x(t) \in \mathbb{R}^N \) is the state of the system at time \( t \), the function \( f: \mathbb{R}^N \rightarrow \mathbb{R}^N \) is a continuously differentiable vector field, and the \( N \times N \) matrix \( W \) is the weight matrix describing the network of the system as illustrated in Figs. 2a and 2b. More specifically, we are interested in a subclass of the latter dynamics with the form \( \dot{x} = g(x,y) \), where \( g: \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^N \) and \( y = Wx \).

Considering this subclass of dynamics already highlight an important implication of the low-rank hypothesis. The linear function \( x \mapsto y = Wx \) in \( g \) has a very special role: even if \( x \) is part of a \( N \)-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 is most likely to have a low effective rank. We can hence say that \( Wx \) will be part of an effectively low-dimension submanifold.

Yet, the vector field \( g \) is a nonlinear function of \( Wx \) and it is not straightforward to assess the low dimensionality of \( g(x,y) \): this is a fundamental challenge of dimension reduction. A considerable number of works have been done on the subject in recent years [34], but it remains unclear how to choose a dimension for the reduced dynamics and how to quantify the corresponding error with the complete dynamics.

The dimension reduction of a dynamical system can be imagined as a problem of aligning a low-dimensional vector field with a high-dimensional vector field (Fig. 2c and SI III A). When perfectly aligned, the dimension reduction is exact; otherwise, there is an alignment error in \( \mathbb{R}^n \) that depends on the position \( x \) in \( \mathbb{R}^N \). Defining a dimension-reduction method requires the choice of a \( n \times N \) reduction matrix \( M \), that maps the elements of the complete system to the reduced system, as well as a vector field \( F \), that describes the evolution of the set of observables \( (X_n)_{n=1}^{\infty} \) in \( \mathbb{R}^n \). The alignment error in \( \mathbb{R}^n \) at \( x \in \mathbb{R}^N \), denoted \( E(x) \), can then be defined as the root mean-square error between the vector fields \( M \circ f \) and \( F \circ M \) (see Methods).

In principle, one would like to minimize the alignment
error to find the best pair \((M, F)\). However, this is a challenging optimization problem in general, even if \(f \) and \(F\) are linear functions (see SI III A). Among all the \(n \times N\) real matrices \(M\), there is not necessarily a unique optimal choice given that this choice depends upon the goal of the modeler. For instance, one could choose \(M\) such that the evolution of \(X\) in time has a clear interpretation for each of its element for all time (e.g., synchronization observables \([34]\)), which may complicate even more the optimization problem. In what follows, we do not impose such restrictions, apart from assuming that \(M\) is a linear transformation, but we find good reasons to choose \(M\) as the \(n\)-truncated matrix of right singular vectors \(V_n^T\) of the weight matrix \(W\).

Let us concentrate on finding an optimal vector field \(F\) without taking into account \(M\) for now. Finding an optimal solution for the alignment error in \(\mathbb{R}^n\) as defined above is typically far from straightforward. However, using the least-squares theorem, we proved that the vector field \(M \circ f \circ M^+\) minimizes an alignment error in \(\mathbb{R}^N\), where \(M^+\) denotes the Moore–Penrose pseudoinverse of \(M\) (see Methods). Doing so allowed us to show, for the general class of dynamics on networks \(\dot{x} = g(x, y)\), that the alignment error \(\mathcal{E}(x)\) caused by the least-square optimal vector field is upper-bounded as

\[
\sqrt{n} \mathcal{E}(x) \leq \|MJ_y'(I - M^+M)x\| + \|W(I - M^+M)\|_2 \|MJ_y'\|_2 \|x\|, \tag{1}
\]

where \(\|\cdot\|_2\) denotes the spectral matrix norm, the element \((i, j)\) of the Jacobian matrices \(J_x = J_x(x', Wx')\) and \(J_y = J_y(x', Wx')\) are the partial derivatives of \(g\) according to \(x\) and \(y\) respectively, with \(x'\) being some point between \(x\) and \(M^+Mx\) (see Methods).

Interestingly, the previous inequality suggests a non-arbitrary way of selecting the reduction matrix. Indeed, \(M = V_n^T\) minimizes the factor \(\|W(I - M^+M)\|_2\), related to the interactions among the elements of the system (see Methods). This choice for \(M\) also has a notable consequence: each observable \(X_\mu\) generally becomes a global observable in that it contains information on most vertices. This characteristic, alongside that it is a finite-size dimension reduction, make our approach stands out from many mean-field modeling 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.

The choice made in Eq. (2) prompted us to derive an-
other inequality revealing the contribution of the network singular values to the alignment error (see Methods, Theorem 2):
\[
\sqrt{n} \mathcal{E}(x) \leq \|V_n^T J'_x (I - P)x\| + \sigma_{n+1} \|V_n^T J'_y\|_2 \|x\|,
\] (3)
where \( P = V_n V_n^\top \). The upper bound is not meant to be as tight as possible, but intuitive. For instance, the inequality provides a criterion for exact dimension reduction. Indeed, if \( J'_x = dI \) for some real constant \( d \) and \( n = r \), the upper bound vanishes to zero and the dimension reduction is exact (recall that \( \sigma_i = 0 \) for \( i > r \); see Methods). As a consequence, a general class of dynamics, including recurrent neural networks and the Wilson-Cowan neuronal dynamics, can be exactly reduced (see Methods). Moreover, the theorem connects the rapid decay of singular values of the networks to the error in the vector field of the reduced dynamics.

Figure 3 shows a rapid decrease of the error with the dimension \( n \), the latter being in accordance with the rapid decay of the upper bound and of the singular values. In SI III D-SI III F, we show how the least-square reduced dynamics satisfactorily predicts the emergence of an epidemic in the epidemiological dynamics or the hysteresis of the global activity (resp. population size) of the neuronal (resp. microbial) dynamics on real networks (Fig. S8 and Fig. S11). Altogether, as schematized in Fig. 2, the low-rank hypothesis induces a low-dimension hypothesis for dynamics on network.

The reduced system is akin to a low-dimensional dynamics taking place on a smaller structure, whose nature remains to be specified (Fig. 2c). We show in the next section that dimension reduction ultimately leads to the emergence of higher-order interactions, as illustrated in Fig. 2d.

**Emergence of higher-order interactions**

Theoretical and experimental evidence of the existence of higher-order interactions in various complex systems have been reported and their consequences—e.g., on explosive transitions [35] or mesoscopic localization [36]—have been extensively studied [37]. However, the origin of these higher-order interactions in complex systems remains under active investigation, notably for oscillatory systems [38, 39] (see SI III C).

Using our framework, a simple example readily provides insights over 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 for the vertex \( i \) to be infected, \( y_i = W x \) while \( d_i \) and \( \gamma \) are constants denoting the recovery rate of vertex \( i \) and the infection rate respectively. The reduced dynamics is then given by
\[
\dot{X}_\mu = \sum_{\nu=1}^n (D_{\mu\nu} + W_{\mu\nu}) X_\nu \quad (4)
\]
\[-\gamma \sum_{i=1}^N M_{\mu i} \left( \sum_{\nu=1}^n M_{\nu i}^T X_\nu \right) \left( \sum_{j=1}^N \sum_{k=1}^n W_{ij} M_{jk}^T X_k \right)
\]
for all \( \mu \in \{1, ..., n\} \), where \( D = -MDM^+ \) is a reduced \( n \times n \) recovery rate matrix with \( D = \text{diag}(d_1, ..., d_N) \), and
$W = \gamma MW M^+ \text{ is a reduced } n \times n \text{ weight matrix.}$

Let us inspect the last term in Eq. (4) more carefully. For the sake of clarity, consider that $M^+ = M^T$, i.e., $M$ has orthogonal rows. Then, $M_{\mu i}$ quantifies the influence of vertex $i$ on the $\mu$-th observable, $M^\mu_i X_\nu$ is the influence of the $\nu$-th observable weighted by its dependence over vertex $i$, and $W_{ij} M_{\nu j} X_\kappa$ is the influence of the $\kappa$-th observable weighted by its dependence over vertex $j$ that connects to vertex $i$. Altogether, these factors thus form a third-order interaction between the observables $X_\mu$, $X_\nu$, and $X_\kappa$, an observation that becomes even more explicit by rearranging Eq. (4) as

\begin{equation}
\dot{X}_\mu = \sum_{\nu=1}^{n} (D_{\mu \nu} + W_{\mu \nu}) X_\nu + \sum_{\nu, \kappa=1}^{n} T_{\mu \nu \kappa} X_\nu X_\kappa,
\end{equation}

where the third-order interactions are encoded in a third-order tensor $T$ with elements

\begin{equation}
T_{\mu \nu \kappa} = -\gamma \sum_{i,j=1}^{N} M_{\mu i} M_{\nu j} W_{ij} M_{\nu j}^\kappa,
\end{equation}

for all $\mu, \nu, \kappa \in \{1,...,n\}$. Hence, the resulting structure of the reduced system is a hypergraph $H$ with $n$ vertices (Fig. 2c-d; see SI III C), which is generally directed [40], weighted, signed, and formed from $D, W$, and $T$.

Apart from the dependencies over the dynamical parameters such as the weight matrix $W$, Eq. (6) highlights the important contribution of the reduction matrix $M$ in the higher-order interaction. Indeed, $M$ partially determines the directed, weighted, and signed nature of the hypergraph. Moreover, if the observables respectively depend on disjoint groups of vertices, i.e., $M_{\mu i} \propto \delta_{\mu s(i)}$, where $\delta$ is the Kronecker delta and the surjection $s$ maps each vertex $i$ to its group, then the tensor with elements in Eq. (6) can be exactly mapped to a matrix. In other words, in the epidemiological dynamics, the higher-order interactions emerge from observables depending on overlapping groups of vertices (e.g., $M = V_n^T$ in general). Interestingly, such overlapping is a very common characteristic of complex networks, notably in social networks [41].

The latter observations encouraged us to seek generic conditions for such emergence. For a general dynamics of the form $\dot{x}_i = h_i(x_i, y_i)$, where $h_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ is an analytical vector field for all $i \in \{1,...,N\}$, we proved that the least-square reduced vector field depends upon higher-order interactions between the observables $X_1,...,X_n$ (see Methods, Proposition 3). We then deduced two insightful consequences. First, if the vector field is a polynomial of total degree $\delta$ in $x_i$ and $y_i$ for all $i$, then the hypergraph of the reduced system has interactions of maximal order $\delta + 1$ (see Corollary S52). 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 (see Corollary S53). Other worked-out examples for a microbial and an oscillator dynamics are given in Table III of the Methods, which complement the previous observations on the epidemiological dynamics.

All in all, 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 to the ensuing interactions of the reduced system.

**Conclusions and outlook**

In this paper, 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 something intrinsic to many real complex systems. 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 doing the low-rank hypothesis unknowingly can lead to oversimplified complex systems’ models.

Experimentally, when observing time series of a complex system at a relatively coarse-grained resolution (e.g., local field potentials in the brain [42] or abundances in plant communities [43]), our theoretical results suggest that it is somewhat expected to find significant higher-order interactions by measuring correlations or by fitting higher-order models to these data. We thus urge experimentalists to take up the challenge of monitoring complex systems at different scales to evaluate the empirical consequences of the dimension at which the measurements are done on the emergence of higher-order interactions.

Nevertheless determining the precise form of the dominant observables that drive the behavior of complex systems remains an open problem. While we focused on linear observables, there might exist a small set of nonlinear observables well suited for a given high-dimensional dynamics [44]. However, finding appropriate, intuitive, nonlinear observables is much harder [45]. We hope that our results will encourage the community to further explore analytically the delicate balance between a sufficiently detailed microscopic description and an interpretable macroscopic one, as well as the emergent features that come as a byproduct of dimension reduction.

Finally, one defining property of complex systems that we have not addressed is their capacity for adaptation [46]. Preliminary results indicate that the differential equations describing adaptation require a special analytical treatment for the dimension-reduction framework to fully grasp the impact of adaptation on the system’s behavior. Our results also suggest that the low-rank hypothesis plays a central role for controlling [47, 48] and assessing the resilience of adaptive complex systems [49]. This, alongside indications that maturation or learning could reduce the effective rank in *C. elegans* (see SI II E) and in a wide range of artificial neural networks [50], will be the topic of an upcoming publication.
[1] F. Y. Kuo and I. H. Sloan, Not. Am. Math. Soc. 52, 1320 (2005).
[2] S. Ganguli and H. Sompolinsky, Annu. Rev. Neurosci. 35, 485 (2012).
[3] L. F. Abbott and al., Cell 182, 1372 (2020).
[4] P. W. Anderson, Science 177, 393 (1972).
[5] S. Strogatz, S. Walker, J. M. Yeomans, C. Tarnita, E. Arcaute, M. De Domenico, O. Artine, and K.-I. Goh, Nat. Rev. Phys. 4, 508 (2022).
[6] R. M. May, Nature 261, 459 (1976).
[7] J. von Neumann, in John von Neumann Collected Work, Vol. V, edited by A. H. Taub (Bergammon Press, 1963) p. 288.
[8] S. Wolfram, Nature 311, 419 (1984).
[9] G. Parisi, Nature 530, 307 (2016); C. Tu, J. Grilli, F. Schuesler, and S. Suweis, Phys. Rev. E 95, 062307 (2017); J. Jiang, Z.-G. Huang, T. P. Seager, W. Lin, C. Grebogi, A. Hastings, and Y.-C. Lai, Proc. Natl. Acad. Sci. U.S.A. 115, E639 (2018); E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, Phys. Rev. X 9, 011042 (2019); V. Thibeault, G. St-Onge, L. J. Dubé, and P. Desrosiers, Phys. Rev. Research 2, 043215 (2020); M. Vegué, V. Thibeault, P. Desrosiers, and A. Allard, arXiv:2206.11230 (2022); P. Kundu, H. Kori, and N. Masuda, Phys. Rev. E 105, 024305 (2022).
[10] S. Fortunato and M. E. J. Newman, Nat. Phys. 18, 848 (2022).
[11] H. Weyl, Math. Ann. 71, 441 (1912).
[12] K. Fan, Proc. Natl. Acad. Sci. U.S.A. 37, 760 (1951).
[13] W. E. Donath and A. J. Hoffman, IBM J. Res. Dev. 15, 550 (1971).
[14] M. Beiran, A. Dubreuil, A. Valente, F. Mastrogiuseppe, and S. Ostojic, Neural Comput. 33, 1572 (2021).
[15] P. Gao and S. Ganguli, Curr. Opin. Neurobiol. 32, 148 (2015).
[16] A. F. K. Leung, P. F. Abrahams, and C. D. Roy, Proc. Natl. Acad. Sci. U.S.A. 119, 213750119 (2022).
[17] B. Beckermann and A. Townsend, SIAM J. Matrix Anal. Appl. 38, 1227 (2017).
[18] M. Udell and A. Townsend, SIAM J. Math. Data Sci. 1, 144 (2019).
[19] J. Gao, B. Barzel, and A.-L. Barabási, Nature 530, 307 (2016); C. Tu, J. Grilli, F. Schuesler, and S. Suweis, Phys. Rev. E 95, 062307 (2017); J. Jiang, Z.-G. Huang, T. P. Seager, W. Lin, C. Grebogi, A. Hastings, and Y.-C. Lai, Proc. Natl. Acad. Sci. U.S.A. 115, E639 (2018); E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, Phys. Rev. X 9, 011042 (2019); V. Thibeault, G. St-Onge, L. J. Dubé, and P. Desrosiers, Phys. Rev. Research 2, 043215 (2020); M. Vegué, V. Thibeault, P. Desrosiers, and A. Allard, arXiv:2206.11230 (2022); P. Kundu, H. Kori, and N. Masuda, Phys. Rev. E 105, 024305 (2022).
[20] E. Arcaute, M. De Domenico, O. Artime, and K.-I. Goh, Phys. Rev. X 9, 031050 (2019).
[21] S. Fortunato and M. E. J. Newman, Nat. Phys. 18, 848 (2022).
[22] S. Fortunato and M. E. J. Newman, Nat. Phys. 18, 848 (2022).
[23] V. A. Marˇ cenko and L. A. Pastur, Math. USSR-Sbornik 1, 457 (1973).
[24] M. Avish and D. L. Donoho, IEEE Trans. Inf. Theory 63, 2137 (2017).
[25] G. Yan, P. E. Vértes, E. K. Towlson, Y. L. Chew, D. S. Walker, W. R. Schafer, and A.-L. Barabási, Nature 550, 519 (2017).
[26] A. Marčenko and L. A. Pastur, Math. USSR-Sbornik 1, 457 (1967).
[27] 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).
[28] D. M. Cvetkovic, M. Doob, and H. Sachs, “Spectra of graphs. Theory and application,” (1980); F. Chung, Spectral Graph Theory (CBMS, Rhode Island, 1994); B. Nica, A Brief Introduction to Spectral Graph Theory (European Mathematical Society, Zurich, 2018).
[29] H. Weyl, Math. Ann. 71, 441 (1912).
[30] K. Fan, Proc. Natl. Acad. Sci. U.S.A. 37, 760 (1951).
[31] E. W. Kuo and I. H. Sloan, Not. Am. Math. Soc. 52, 1320 (2005).
[32] S. Ganguli and H. Sompolinsky, Annu. Rev. Neurosci. 35, 485 (2012).
[33] L. F. Abbott and al., Cell 182, 1372 (2020).
[34] P. W. Anderson, Science 177, 393 (1972).
[35] S. Strogatz, S. Walker, J. M. Yeomans, C. Tarnita, E. Arcaute, M. De Domenico, O. Artine, and K.-I. Goh, Nat. Rev. Phys. 4, 508 (2022).
[36] R. M. May, Nature 261, 459 (1976).
[37] J. von Neumann, in John von Neumann Collected Work, Vol. V, edited by A. H. Taub (Bergammon Press, 1963) p. 288.
[38] S. Wolfram, Nature 311, 419 (1984).
[39] G. Parisi, Nature 530, 307 (2016); C. Tu, J. Grilli, F. Schuesler, and S. Suweis, Phys. Rev. E 95, 062307 (2017); J. Jiang, Z.-G. Huang, T. P. Seager, W. Lin, C. Grebogi, A. Hastings, and Y.-C. Lai, Proc. Natl. Acad. Sci. U.S.A. 115, E639 (2018); E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, Phys. Rev. X 9, 011042 (2019); V. Thibeault, G. St-Onge, L. J. Dubé, and P. Desrosiers, Phys. Rev. Research 2, 043215 (2020); M. Vegué, V. Thibeault, P. Desrosiers, and A. Allard, arXiv:2206.11230 (2022); P. Kundu, H. Kori, and N. Masuda, Phys. Rev. E 105, 024305 (2022).
[40] E. Arcaute, M. De Domenico, O. Artime, and K.-I. Goh, Phys. Rev. X 9, 031050 (2019).
Methods

Random graphs. A random network can be described by a random matrix
\[ W = \langle W \rangle + X, \]
where \( \langle W \rangle \) is the expected weight matrix and \( X \) 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 \( L \), i.e.,
\[ \langle W \rangle = \left\{ \phi(L_{ij}) \right\}_{i,j=1}^N, \]
where \( \phi \) is a real-valued function of a real variable. In Table I, we list some classical examples of random graphs and the corresponding low-rank matrices.

TABLE I: Low-rank matrix \( L \) characterizing the expected adjacency matrix for different random graphs of \( N \) vertices. SBM: Stochastic Block Model, RDPG: Random Dot Product Graph, RGM: Random Geometric Model. The parameters \( q, r, \) and \( d \) are usually assumed to be small. More details about these random graphs and others are given in SI 1C.

| Model          | Low-rank matrix \( L \) | rank(\( L \)) | \( \phi(L_{ij}) \) |
|----------------|-------------------------|--------------|---------------------|
| \( G(N, p) \)  | \( N_p 11^T \)           | 1            | \( L_{ij} \)        |
| SBM            | \( \sum_{\mu, \nu=1}^q \sqrt{p_{\mu p_{\nu}} b_{\mu} b_{\nu}} \) | \( q \) | \( L_{ij} \)        |
| Chung-Lu       | \( \frac{|x_i|^2}{2d} \) \( R \) \( \tau \) | 1            | \( L_{ij} \)        |
| Metadegree     | \( \sum_{\mu=1}^r \Delta_{\mu \nu} \sqrt{v_{\mu} v_{\nu}^T} \) | \( r \) | \( L_{ij} \)        |
| RDPG           | \( \sum_{\mu=1}^d X_{\mu} \) \( X_{\mu}^T \) | \( d \) | \( L_{ij} \)        |
| DSCM           | \( \alpha \beta \) \( \sigma \) | 1            | \( \frac{L_{ij}}{1+L_{ij}} \) |
| \( S^2 \) RGM   | \( \frac{R}{p} \) \( \frac{(k \tau)}{\sigma} \) \( \circ \) \( \theta \) | \( \leq 3 \) | \( \frac{1}{1+L_{ij}} \) |

It is straightforward to assess the low rank of \( L \), but it is harder to assess the low rank of \( \langle W \rangle \). For the directed self-configuration model (DSCM), we show in Lemma S30 (SI II C) that \( \langle W \rangle \) is an infinite sum of rank 1 matrices. This observation then leads to Theorem 1 which demonstrates that the singular values of \( \langle W \rangle \), in at least two disjoint density regimes, are bounded above by an exponentially decreasing term.

Theorem 1 (Simplified version of Theorem S34). Let \( \sigma_1 \geq \ldots \geq \sigma_N \) be the singular values of the DSCM expected matrix \( \langle W \rangle \).

1. If \( (W_{ij}) < \gamma(1 + \gamma) \) for all \( i, j \in \{1, \ldots, N\} \) and for some \( \gamma \in (0, 1) \), then the singular values of \( \langle W \rangle \) are upper bounded as
   \[ \sigma_i \leq \frac{N \gamma^i}{1 - \gamma}, \quad \forall \ i \in \{1, \ldots, N\}. \]

2. If \( (W_{ij}) > \omega(1 + \omega) \) for all \( i, j \in \{1, \ldots, N\} \) and for some \( \omega > 1 \), then the singular values of \( \langle W \rangle \) are upper bounded as
   \[ \sigma_i \leq N \delta_{i1} + \frac{N \omega^i - 1}{\omega - 1}, \quad \forall \ i \in \{1, \ldots, N\}. \]

The proof is based on Weyl’s inequalities (Theorem S8 in SI 1B) and the truncated geometric series. The upper bounds are not meant to be tight for all \( \langle W \rangle \) and \( \gamma \), but they are sufficient to establish the validity of the low-rank hypothesis in the DSCM. Theorem 1 paves the way for the verification of the low-rank hypothesis in other random graphs, such as the random geometric model (RGM).

Effective ranks. The idea of extracting the number of significant components in a matrix decomposition is an old theme (e.g., in factor analysis [1, 2] or PCA [3, How Many Components ?]), but is still subject to new interesting developments in random matrix theory, data science [4, 5] and in network science where hyperbolic geometry [6] and information theory [7] are used. Because of the close relationship of SVD with the rank, many effective ranks are defined using the singular values. Intuitively, these effective ranks are numbers that indicate how many singular values are significant when decomposing a matrix. Table II presents the list of different effective ranks that we have inventoried.

The effective ranks shrink and shrink are defined from matrix denoising techniques such as the ones introduced by Refs. [4, 5, 8], which rely on spectral theory of infinite random matrices [9] to determine optimal ways of shrinking the singular values distribution (see SI 1C).

| Table II: Different effective ranks of a matrix of dimension \( N \times N \) and of rank \( r \) expressed in terms of its singular values \( \sigma_1 \geq \ldots \geq \sigma_N \). For energy, the constant \( \tau \) is a threshold to be set between 0 and 1. For shrink, \( \text{erank} \) is the median singular value and \( \mu_{\text{med}} \) is the median of a Marcenko-Pastur probability density function \([4]\). For shrink, \( s^{\ast} \) denotes an optimal singular value shrinkage function \([5]\). The complete names and the details about each of the effective ranks are given in SI 1C. |

| Model          | Expression                                           |
|----------------|------------------------------------------------------|
| rank           | \( \sum_{i=1}^r \sigma_i^2 / \sigma_i \)            |
| nrank          | \( \sum_{i=1}^r \sigma_i / \sigma_i \)              |
| energy         | \( \min \left[ \arg \max_{x_{\{1, \ldots, N\}}} \left( \frac{x_{\{r\}}}{\sqrt{\sum_{j=1}^r x_{\{j\}}}} \right)^2 \right] > \tau \) |
| elbow          | \( \arg \max_{x_{\{1, \ldots, N\}}} \left( \frac{\sum_{j=1}^r x_{\{j\}}}{\sqrt{\sum_{j=1}^r x_{\{j\}}}} \right)^2 \) |
| shrink         | \# \( \{ \sigma_i \} \) \( i \in \{1, \ldots, N\} \) and \( \sigma_i > \frac{\mu_{\text{med}}}{\sqrt{\nu_{\text{med}}}} \) |
| shrink         | \# \{ \text{st}(\sigma_i) \} \( i \in \{1, \ldots, N\} \) and \( \text{st}(\sigma_i) > 0 \) |

Dimension reduction of dynamical systems. Dimension reduction of high-dimensional nonlinear dynamics is a powerful technique to get analytical and numerical insights on complex systems. Low-dimensional dynamics can be obtained from an optimization problem, where 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 \( F \) 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. In SI III B, we provide a useful diagram (see Diagram S114) that sheds light on the links between the different ways to define alignment errors between vector fields.

More precisely, 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 \( M^+ \circ F \circ f \).

\[ \varepsilon(x) = \| f(x) - M^+ \circ F(M) \| / \sqrt{N}; \]

and the alignment error in \( \mathbb{R}^n \) is the RMSE between the vector field \( M^+ \circ f \circ M \).

\[ \varepsilon(x) = \| M^+ f(x) - F(M) \| / \sqrt{N}; \]

where \( \| \cdot \| \) 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 also define the alignment errors

\[ \varepsilon'(x) = \| f(P) - M^+ \circ F(M) \| / \sqrt{N}; \]

\[ \varepsilon''(x) = \| M f(P) - F(M) \| / \sqrt{n}. \]
with $P = M^+ M$ being a projector and $M^+$ being the Moore–Penrose pseudoinverse of $M$. In principle, the alignment error $E(x)$ in $\mathbb{R}^n$ is to be minimized in order to be as close as possible to an exact dimension reduction (see Definition S35, Theorem S36, and Diagram S108), but this is far from a simple task. However, as shown in Theorem S39, one can use least squares to show that the vector field of the reduced dynamics

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

(15)

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

In Table III, we carry out the optimal dimension reduction on five dynamics from different fields of application. For the RNN and the neuronal dynamics, we have $D(\delta) = -MDM^+$ where $D = \text{diag}(d_1, \ldots, d_N)$ and $\mu_{ij} = \sum_{k=1}^N W_{ik} M_{kj}$ and we discuss about the other dynamics in the next part of the Methods. With the optimal vector field in Eq. (15), and for dynamics of the general form $\dot{x} = g(x, y)$ (see Assumptions S57), we find an upper bound on the alignment error $E(x)$ related to the singular values of $W$.

Theorem 2 (Simplified version of Theorem S59). The alignment error $\mathcal{E}(x)$ in $\mathbb{R}^n$ at $x \in \mathbb{R}^N$ is upper-bounded as

$$\sqrt{\mathcal{E}(x)} \leq \left\| V_{i} (I - V_{i} V_{i}^T) x + \sigma_{n+1} \left\| V_{i} J_{i}^r \right\|_2 \right\|_2,$$

(16)

where $y' = W x'$ with $x'$ being some point between $x$ and $V_{i} V_{i}^T x$, $\sigma_{i}$ is the $i$-th singular value of $W$, and $J_{i} = J_{i}(x', y')$. $J_{i}$ is the Jacobian matrix of $f$ 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 holds:

$$\mathcal{E}(x) \leq \frac{1}{\sqrt{n}} \left( \sigma_{n+1} + \sigma_{n+2} \right),$$

(17)

where $\sigma_{n+1} = \sigma_{1}(J_{n+1}(x', y'))$ and $\sigma_{n+2} = \sigma_{1}(J_{n+2}(x', y'))$.

As a bonus, the proof of the theorem suggests choosing $M$ as the truncated right singular functions $V_{i}^r$, since it allows minimizing a part of the bound. This is a consequence of the Schmidt-Eckart-Young-Mirsky theorem and more specifically, Theorem S11. Theorem 2 also provides a criterion for exact dimension reduction: if $J_{n}(x', y') = df$ for some real constant $d$ and $n = \text{rank } W$, then $\mathcal{E}(x) = 0$ (see Corollary S61 in SI III D). For example, we find that the class of dynamics of matrix form

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

(18)

where $s$ is a vector of $N$ functions $s_{j} : \mathbb{R} \to \mathbb{R}$ and $W$ has rank $r$ and compact SVD $U_{r} \Sigma_{r} V_{r}^T$, can be exactly reduced to the $r$-dimensional reduced dynamics

$$\dot{X} = d X + V_{r}^T s(U_{r} \Sigma_{r} X),$$

(19)

where $X = V_{r}^T x$. For any $n$ and $X = V_{n} x$, the vector-field in Eq. (19) is the least-square optimal one in the sense described in Theorem S39 of the SI III B. This result implies that any RNN or any neuronal dynamics (with $\delta = 1$) is optimal in the sense that it minimizes the alignment error $\mathcal{E}(x)$ for maximal order $\delta + 1$ (Corollary S52). 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 doesn’t hold in general for nonlinear dependencies of $h_{i}$ over $y_{i}$ (Corollary S53).

In Table III, we apply Proposition 3 and Corollary S52 to the QMF SIS dynamics, the microbial dynamics, and the Kuramoto-Sakaguchi dynamics, which illustrates concretely the emergence of higher-order interactions through dimension reduction. More details are given in SI III C.

Data availability. Most of the real networks data used in the paper are available on Netzschleuder. The artificial networks are from the repository NWS [10] and the Mendeley data of the paper by Hadjibabadi et al. [11]. The C. elegans signed network is obtained by completing (with Dale’s principle) the connectome NT+R method prediction of the open-source database Elegan-Sign [12] [see graphs/get_connectome_weight_matrix on the GitHub repository-low-rank-hypothesis-complex-systems]. The mouse connectome at the level of voxels is available in Mendeley data mouse_connectome_voxelwise [13] and the mesoscopic mouse connectome is given in Ref. [14]. The drosophila connectome is taken from Ref. [15]. The zebrafish mesoscopic connectome is adapted from Ref. [16] and the treatment is available on the paper’s GitHub repository low-rank-hypothesis-complex-systems. The human gut microbiome is from Ref. [17] and was constructed as in the supplementary material of Ref. [18]. The majority of the economic networks are from Ref. [19] and from Ref. [20] on Dryad.

Code availability. The Python code used to generate the results of the paper is available on the GitHub repository low-rank-hypothesis-complex-systems. The code for the optimal shrinkage of singular values is a Python implementation of the Matlab codes optimal_singval_threshold [4] and optimal_singval_shrink [5], which is partly based on the repository opfbvl by B. Erickson.

Acknowledgments. We are grateful to Gabriel Eilerstein for sharing the code to extract the weight matrices from the repository NWS, Charles Murphy for useful discussions on machine learning, and Guillaume St-Onge for his comments on the preprint. We thank Émile Boran for his fundamental contribution to linear algebra. This work was supported by the Fonds de recherche du Québec – Nature et technologies (V.T., P.D.), the Natural Sciences and Engineering Research Council of Canada (V.T., A.A., P.D.), and the Sentinelle Nord program of Université Laval, funded by the Canada First Research Excellence Fund (V.T., A.A., P.D.).

Author contributions. All authors contributed to the formulation of the study, the interpretation of the results, and the edition of the paper. V.T. and P.D. obtained the mathematical results and conceived the conceptual basis of the project. V.T. drafted the manuscript, the supplementary information, and wrote the code to generate the results. A.A. assisted V.T. to write the code and perform the numerical experiments related to Fig. 1.

Competing interests. The authors declare no competing interests.
The sigmoid function is denoted as \(\sigma(x)\). The acronym RNN stands for recurrent neural network. The neuronal dynamics is the Wilson-Cowan dynamics, where \(S\) with the index \(i\) and \(\bar{S}\) denotes complex conjugation. This also implies that \(X_1, ..., X_n\) are complex for the Kuramoto-Sakaguchi dynamics. The oscillator dynamics is the Wilson-Cowan dynamics, where \(S\) denotes the sigmoid function \(x \mapsto 1/(1 + e^{-x})\). More details about the dynamics and their parameters are given in SI III B and SI III C.

| Name          | Complete vector field \(h_i(x_i, y_i)\)                                                                 | Reduced vector field \(H_\mu(X_1, ..., X_n)\)                                                                 |
|---------------|--------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Epidemiological | \(-d_i x_i + \gamma (1 - x_i) y_i\)                                                                  | \(\sum_{\nu=1}^n (D_{\mu\nu} + W_{\mu\nu}) X_\nu + \sum_{\nu,\kappa=1}^n T_{\mu\nu\kappa} X_\nu X_\kappa\) |
| Microbial     | \(a - d_i x_i + bx_i^2 - cx_i^3 + \gamma x_i y_i\) \(G_\mu + \sum_{\nu=1}^n D_{\mu\nu} X_\nu + \sum_{\nu,\kappa=1}^n (D_{\mu(\nu,\kappa)} + T_{\mu\nu\kappa}) X_\nu X_\kappa + \sum_{\nu,\kappa,\tau=1}^n D_{\mu(\nu,\kappa,\tau)} X_\nu X_\kappa X_\tau\) |
| Oscillator    | \(i\omega_i x_i + \gamma e^{-i\omega} y_i - \gamma e^{i\omega} x_i^2 \bar{y}_i\)                   | \(\sum_{\nu=1}^n (D_{\mu\nu} + W_{\mu\nu}) X_\nu + \sum_{\nu,\kappa=1}^n T_{\mu(\nu,\kappa)} X_\nu X_\kappa + \sum_{\nu,\kappa,\tau=1}^n D_{\mu(\nu,\kappa,\tau)} X_\nu X_\kappa X_\tau\) |
| RNN           | \(-d_i x_i + \tanh(\gamma y_i + c_i)\)                                                              | \(\sum_{\nu=1}^n D_{\mu\nu} X_\nu + \sum_{i=1}^N M_{\mu i} \tanh\left(\gamma \sum_{\nu=1}^n \nu_{j\nu} X_\nu + c_i\right)\) |
| Neuronal      | \(-d_i x_i + (1 - ax_i) S[b(\gamma y_i - c_i)]\) \(\sum_{\nu=1}^n D_{\mu\nu} X_\nu + \sum_{j=1}^N M_{\mu j} \left(1 - a \sum_{\nu=1}^n \nu_j X_\nu\right) S\left[b \left(\gamma \sum_{\kappa=1}^n \nu_{j\kappa} X_\kappa - c_i\right)\right]\) |

[1] E. R. Malinowski, *Anal. Chem.* **49**, 606 (1977).
[2] E. Sánchez and B. R. Kowalski, *Anal. Chem.* **58**, 496 (1986).
[3] H. Abdi and L. J. Williams, *WIREs Comput. Stat.* **2**, 433 (2010).
[4] M. Gavish and D. L. Donoho, *IEEE Trans. Inf. Theory* **60**, 5040 (2014).
[5] M. Gavish and D. L. Donoho, *IEEE Trans. Inf. Theory* **63**, 2137 (2017).
[6] P. Almagro, M. Boguna, and M. A. Serrano, *arXiv:2110.14507* (2021).
[7] C. W. Lynn and D. S. Bassett, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023473118 (2021).
[8] P. O. Perry, *Cross-Validation for Unsupervised Learning*, Ph.D. thesis, Stanford University (2009).
[9] F. Benaych-Georges and R. R. Nadakuditi, *J. Multivar. Anal.* **111**, 120 (2012).
[10] G. Eilertsen, D. Jönsson, T. Ropinski, J. Unger, and A. Ynnerman, Proceedings of the European Conference on Artificial Intelligence (ECAI 2020) **325**, 1119 (2020).
[11] 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, *Neuron* **109**, 2556 (2021).
[12] B. G. Feinyes, G. S. Szilágyi, Z. Vassy, C. Sőti, and P. Csermely, *PLOS Comput. Biol.* **16**, e1007974 (2020).
[13] L. Coletta, M. Pagani, J. D. Whitesell, J. A. Harris, B. Bernhardt, and A. Gozzi, *Sci. Adv.* **6**, eabb7187 (2020).
[14] S. W. Oh and al., *Nature* **508**, 207 (2014).
[15] L. K. Scheffer et al., *eLife* **9**, 1 (2020).
[16] M. Kunst, E. Laurell, N. Mokayas, A. Kramer, F. Kubo, A. M. Fernandes, D. Förster, M. Dal Maschio, and H. Baier, *Neuron* **103**, 21 (2019).
[17] R. Lim, J. J. T. Cabathat, T. L. P. Martin, H. Kim, S. Kim, J. Sung, C. M. Ghim, and P. J. Kim, *Sci. Data* **7**, 1 (2020).
[18] H. Sanhedrai, J. Gao, A. Bashan, M. Schwartz, S. Havlin, and B. Barzel, *Nat. Phys.* **18**, 338 (2022).
[19] J. Wachs, M. Fazekas, and J. Kertész, *Int. J. Data Sci. Anal.* **12**, 45 (2021).
[20] S. Ranganathan, M. Kivelä, and J. Kanniainen, *PLOS ONE* **13**, e0198807 (2018).
The low-rank hypothesis of complex systems
— Supplementary information —

CONTENTS

I. Preliminaries on singular value decomposition 1
   A. Definition of SVD and its link to the rank 1
   B. Weyl’s theorem and optimal low-rank factorization 2
   C. Effective ranks 5

II. SVD in the study of complex systems 8
   A. Stable rank of graphs 8
   B. SVD of random graphs 10
   C. Exponential decrease of singular values in the directed soft configuration model 13
   D. Directed network centrality measures 18
   E. Adaptative networks 19
   F. Towards dynamical systems 20

III. Dimension reduction of complex systems 20
   A. Notation and generalities on dimension reduction 20
   B. Least-square optimal vector field 23
   C. Emergence of higher-order interactions 25
   D. Upper bound on the alignment error and exact dimension reduction 31
   E. Computation of the upper bound on the alignment error 34
   F. Global bifurcation diagram 38
   G. Numerical efficiency 39

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) [6–10], Karhunen-Loève Transform (KLT) [11–13], Proper Orthogonal Decomposition (POD) [14–16], and Empirical Orthogonal Function (EOF) [17, 18]. In machine learning, some autoencoders have been shown to be at best equivalent to SVD [19, 20]. Even if the subject is old in itself, there are still many interesting developments about SVD, notably in random matrix theory [21–27] where the singular value distribution is often called the eigenvalue distribution of the Wishart, chiral or Laguerre matrix ensembles [28, Chap. 3] or of sample covariance matrices [21, 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 $A^\dagger A$ ($^\dagger$ 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 S4. 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

\[
A = U\Sigma V^\dagger
\]

where \( U = (u_1, \ldots, u_m) \) and \( V = (v_1, \ldots, v_n) \) 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 \( \Sigma \) is a rectangular diagonal matrix of size \( m \times n \) defined as

\[
\Sigma = \begin{pmatrix}
\sigma_1 & 0 & \cdots \\
0 & \sigma_2 & \cdots \\
\vdots & \ddots & \ddots
\end{pmatrix}
\]

with \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \) and \( \sigma_{r+1} = \cdots = \sigma_q = 0 \)

where \( q = \min(m, n) \) and \( \sigma_i = \sqrt{\lambda_i} \) with \( \lambda_i \) 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 [29], theorem 2.6.3 of Ref. [30], or theorem 1.3.9 of Ref. [23]. \( \square \)

Remark S5. The nonnegative numbers \( \sigma_1, \ldots, \sigma_q \) in the previous theorem are called the singular values of \( A \) while the vectors \( u_1, \ldots, u_m \) and \( v_1, \ldots, v_n \) 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 \( \sigma_i \) as a function of \( A \) and write its values as \( \sigma_i(A) \).

Remark S6. 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 [30, p.151]). Its relation to dimension reduction then becomes obvious: one can truncate the matrices \( U, V, \Sigma \) by removing their last columns (and rows for \( \Sigma \)) to get smaller matrices \( U_r = (u_1 \ldots u_r), V_r = (v_1 \ldots v_r) \), and \( \Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r) \) with \( r = \text{rank}\, A \), and obtain a rank factorization:

\[
A = U_r\Sigma_r V_r^\dagger,
\]

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

Remark S7. It is often more convenient to rewrite the SVD in Eq. (S1) or equivalently in Eq. (S3) as

\[
A = \sum_{i=1}^{r} \sigma_i u_i v_i^\dagger.
\]

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 \( A \),

\[
A^+ = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^\dagger,
\]

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 [30], 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 [31, Theorem 2].

Theorem S8. Let \( A \) and \( B \) be two matrices of dimension \( m \times n \) and let \( q = \min(m, n) \). Then,

\[
\sigma_{i+j-1}(A + B) \leq \sigma_i(A) + \sigma_j(B) \quad \forall \, 1 \leq i, j, \, i + j - 1 \leq q,
\]

(S6)
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 [30]. A proof that uses the Courant-Fisher theorem for singular values is also given in Ref. [29, Theorem 3.3.16].

**Remark S9.** 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 [32] to Wittmeyer [33, Eq. (VIII)], but Wittmeyer himself writes in a footnote that the equation is in Wintner, “Spektraltheorie der unendlichen Matrizen”, Leipzig 1929, p. 130. Nowadays, the result is, perhaps, not surprising: it is the triangle inequality for the spectral matrix norm.

The importance of the Weyl theorem in the paper relies partly on what it implies for dimension reduction. In particular, it allows proving the Schmidt-Eckart-Young-Mirsky theorem [1–3, 34–36] (often called the Eckart-Young inequality for the spectral matrix norm.

For our paper, especially to find the upper bound on the alignment error [Theorem S59], Theorem S10 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 [39, 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 S59.

**Theorem S11.** Let \( A \) be a matrix of rank greater than or equal to \( k \). Consider the optimization problem

\[
\begin{aligned}
\text{minimize} & \quad \| A - B \|_F^2 \\
\text{subject to} & \quad \text{rank } B \leq k,
\end{aligned} \tag{P0}
\]

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

\[
\min_{B, \text{rank } B \leq k} \| A - B \|_2^2 = \sigma_{k+1}^2 \quad \text{or} \quad \min_{B, \text{rank } B \leq k} \| A - B \|_F^2 = \sum_{i=k+1}^q \sigma_i^2, \tag{S8}
\]

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.,

\[
B^* = \arg \min_{B, \text{rank } B \leq k} \| A - B \|_2^2 = \arg \min_{B, \text{rank } B \leq k} \| A - B \|_F^2 = \sum_{i=1}^k \sigma_i u_i v_i^\top, \tag{S9}
\]

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 S59], Theorem S10 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 [39, 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 S59.

**Theorem S11.** Let \( A \) be a matrix of rank \( r \) with singular value decomposition \( A = U \Sigma V^\top \) and \( k \)-truncated singular value decomposition \( U_k \Sigma_k V_k^\top \). Let \( \| \cdot \| \) denote the spectral norm \( \| \cdot \|_2 \) or the Frobenius norm \( \| \cdot \|_F \). Consider the optimization problem

\[
\begin{aligned}
\text{minimize} & \quad \| (I - M^\top M) A \|_F^2, \quad \tag{P1}
\end{aligned}
\]

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

\[
\min_{M, \text{rank } M \leq k} \| (I - M^+ M) A \|_2^2 = \sigma_{k+1}^2 \quad \text{and} \quad \min_{M, \text{rank } M \leq k} \| (I - M^+ M) A \|_F^2 = \sum_{i=k+1}^{\min(m, n)} \sigma_i^2, \tag{S10}
\]

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

\[
\min_{M, \text{rank } M \leq k} \| B(I - M^+ M) \|_2^2 \tag{P2}
\]

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 \( M = R_k^\ell \) minimizes problem (P2) with errors

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

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

\[
\| (I - M^+ M) A \|_2^2 = \| A - AA^+ A \|_2^2 = \| A - A \|_2^2 = 0, \tag{S12}
\]

since \((A^+) + = A\) and \( AA^+ A = A\) by the defining properties of the Moore-Penrose pseudo-inverse [40].

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 \( \| (I - M^+ M) A \|_2^2 = 0 \).

3. We first prove that

\[
\min_{M, \text{rank } M \leq k} \| (I - M^+ M) A \|_2^2 = \min_{C} \| A - C \|_2^2. \tag{S13}
\]

Indeed, due to Sylvester’s rank inequality [30, Section 0.4.5 (c)] and the inequality \( \text{rank } M \leq k \),

\[
\text{rank}(M^+ M) \leq \min\{\text{rank } M, \text{rank } M^+\} \leq k, \tag{S14}
\]

which in turn implies that

\[
\text{rank}(M^+ M A) \leq \min\{\text{rank}(M^+ M), r\} = \text{rank}(M^+ M) \leq k, \tag{S15}
\]

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

\[
\min_{M, \text{rank } M \leq k} \| (I - M^+ M) A \|_2^2 = \min_{C} \| A - C \|_2^2. \tag{S16}
\]

Let us now focus on the Frobenius norm. The new form of the problem in Eq. (S16) is compatible with Theorem S10, but with the additional equality constraint that \( C = M^+ M A \), which directly implies the inequality

\[
\min_{C} \| A - C \|_2^2 \geq \min_{\text{rank } C \leq k} \| A - C \|_2^2 = \sum_{i=k+1}^{\min(m, n)} \sigma_i^2, \tag{S17}
\]

or equivalently, from Eq. (S16),

\[
\min_{M, \text{rank } M \leq k} \| (I - M^+ M) A \|_F^2 \geq \sum_{i=k+1}^{\min(m, n)} \sigma_i^2. \tag{S18}
\]

Therefore, if we find a matrix \( M \) that reaches the lower bound of inequality (S18), 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,

\[
(I - M^+ M)^\top (I - M^+ M) = (I - M^+ M)^2 = I - M^+ M. \tag{S19}
\]
The cyclic property of the trace and the eigenvalue decomposition of $AA^\top$ from the SVD imply
\[
\| (I - M^+ M) A \|_F^2 = \text{tr} \left[ AA^\top (I - M^+ M) \right] = \text{tr} \left[ U \Sigma^2 U^\top (I - M^+ M) \right].
\]  
(S20)

Let $M$ be equal to $U_k^\top$. Then, $M^+ M = U_k U_k^\top$ and
\[
\| (I - M^+ M) A \|_F^2 = \text{tr} \left[ \sum_{i=1}^{\min(m,n)} \sigma_i^2 u_i u_i^\top - \sum_{i=1}^{\min(m,n)} \sum_{j=1}^{k} \sigma_i^2 u_i u_i^\top u_j u_j^\top \right].
\]  
(S21)

Since $u_i^\top u_j = \delta_{ij}$, we obtain
\[
\sum_{i=1}^{\min(m,n)} \sum_{j=1}^{k} \sigma_i^2 u_i u_i^\top u_j u_j^\top = \sum_{i=1}^{\min(m,n)} \sum_{j=1}^{k} \sigma_i^2 u_i \delta_{ij} u_j^\top = \sum_{j=1}^{k} \sigma_j^2 u_j u_j^\top
\]  
(S22)

and thus,
\[
\| (I - M^+ M) A \|_F^2 = \text{tr} \left[ \sum_{i=k+1}^{\min(m,n)} \sigma_i^2 u_i u_i^\top \right] = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2 = \sum_{i=r+1}^{\min(m,n)} \sigma_i^2
\]  
(S23)

Hence, $B^* = U_k U_k^\top A$ 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 $(I - M^+ M) A = (I - U_k U_k^\top) U \Sigma V^\top$ which is equal to
\[
\sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^\top - \sum_{i=1}^{k} \sigma_i \sum_{j=1}^{k} u_j u_j^\top u_i v_i^\top - \sum_{i=k+1}^{\min(m,n)} \sigma_i \sum_{j=1}^{k} u_j u_j^\top u_i v_i^\top = \sum_{i=k+1}^{\min(m,n)} \sigma_i u_i v_i^\top,
\]  
(S24)

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 S11 as explained in the following example.

**Example S12.** 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 $U_n = (u_1 \ldots u_n)$ with $u_\mu$ being the $\mu$-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 [36, 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 [41, Definition 7.6.7], also called numerical rank [42], is defined as
\[
srank(A) = \frac{\| A \|_F^2}{\| A \|_2^2} = \frac{\sum_{i=1}^{r} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2}.
\]  
(S25)

It thus measures of 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
\[
\| A \|_2 \leq \| A \|_F \leq \sqrt{r} \| A \|_2,
\]  
(S26)

where $r = \text{rank}(A)$. From the second inequality, we easily deduce the following upper bound:
\[
srank(A) \leq r.
\]  
(S27)

The stable rank is stable in the sense that it remains essentially unchanged under a small perturbation of the matrix $A$, contrary to the rank [42]. It is used in the design of fast (randomized) algorithms for low-rank approximations [39, 43]. 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 [44].
The nuclear rank \cite[p.2183]{45} is defined as
\[
\text{nrank}(A) = \|A\|_* = \frac{\sum_{i=1}^r \sigma_i}{\sigma_1},
\] (S28)
where \(\|\cdot\|_*\) 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
\[
\|A\|_* \leq \sqrt{r} \|A\|_F \leq r \|A\|_2,
\] (S29)
where we have used the first inequality of Eq. (S26). Therefore, we find that
\[
\text{nrank}(A) \leq \sqrt{r} \text{erank}^{1/2}(A) \leq r.
\] (S30)

- The energy ratio, also called the cumulative explained variance, the reconstructed proportion, or the \(R_v\) coefficient \cite{10}, is
\[
E(\ell) = \frac{\|A_{\ell}\|_F^2}{\|A\|_F^2} = \frac{\sum_{i=1}^\ell \sigma_i^2}{\sum_{j=1}^r \sigma_j^2},
\] (S31)
where \(A_{\ell}\) is the \(\ell\)-truncated SVD of \(A\). The energy ratio effective rank is
\[
\text{energy}(A) = \min \left\{ \frac{1}{\ell \in \{1, \ldots, N\}} \left( E(\ell) > \tau \right) \right\},
\] (S32)
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 \cite{46–48}.
- Let the coordinate \((x_i, y_i)\) of the \(i\)-th singular values be given by
\[
x_i = \frac{i - 1}{N - 1} \quad \text{and} \quad y_i = \frac{\sigma_i - \sigma_N}{\sigma_1 - \sigma_N},
\] (S33)
for all \(i \in \{1, \ldots, N\}\), such that the largest singular value is at \((0, 1)\) and the smallest singular value is at \((1, 0)\). The distance between the line \(L = \{(x, y)\mid x + y = 1\}\), passing through the largest and the smallest singular value, and the position \((x_i, y_i)\) of the \(i\)-th singular value is
\[
d_i = \frac{1}{\sqrt{2}} |x_i + y_i - 1|.
\] (S34)
The elbow position is the largest distance between in \(\{d_1, \ldots, d_N\}\), i.e.,
\[
i_{\text{elbow}} = \arg \max_{i \in \{1, \ldots, N\}} d_i.
\] (S35)
The elbow rank is thus defined as the number of singular values above the position of the elbow, which is described by
\[
\text{elbow}(A) = i_{\text{elbow}} - 1 = \frac{1}{\sqrt{2}} \arg \max_{i \in \{1, \ldots, N\}} \left| \frac{i - 1}{N - 1} + \frac{\sigma_i - \sigma_N}{\sigma_1 - \sigma_N} - 1 \right| - 1.
\] (S36)
This effective rank is often used as a rule of thumb to truncate the singular value distribution \cite{25, 49}. It is also named the “scree” or elbow test \cite{10} and may be computed in different ways than above \cite{9}.
- Roy and Vetterli’s effective rank \cite{50} or Cangelosi and Goriely’s information dimension \cite{51} is here called erank. It is defined as
\[
\text{erank}(A) = \exp[H(p_1, \ldots, p_r)]
\] (S37)
where \(H(p_1, \ldots, p_r) = -\sum_{i=1}^r p_i \log p_i\) is the Shannon (spectral) entropy, measured in nat, as a function of the singular value mass function
\[
p_i = \frac{\sigma_i}{\|A\|_*} = \frac{\sigma_i}{\sum_{j=1}^r \sigma_j}, \quad \forall i \in \{1, \ldots, r\}.
\] (S38)
Note that the square of the singular values could be used to define the singular value mass function, as in Ref. \cite{52}. Among other interesting properties, the erank satisfies \(1 \leq \text{erank}(A) \leq r\) and it is naturally related to the minimum coefficient rate \cite{53} (see Ref. \cite[Sec. 3]{50} for more details).
- Let \(A = A_{\ell} + X\) where \(A_{\ell}\) is a (deterministic) matrix of unknown rank \(\ell\) and \(X\) is some noise random matrix.
Based on Ref. [54, Definition 4.2] and Ref. [25], the **optimal threshold** $\tau^*(A)$ is defined as

$$
\tau^*(A) = \arg \min_{\tau} \| A_\ell - \hat{A}(\tau) \|, \tag{S39}
$$

where $\hat{A}(\tau)$ is the $\tau$-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-$\ell$ matrix $A_\ell$ (signal matrix) by removing the “noisy” singular values of $A$ due to $\gamma X$. When the level of noise is unknown, under some conditions on $X$, the optimal threshold

$$
\tau^*(A) = \frac{4\sigma_{\text{med}}}{\sqrt{3}\mu_{\text{med}}}, \tag{S40}
$$

minimizes the Frobenius norm $\| A_\ell - \hat{A}(\tau) \|_F$ in the limit of infinite matrices [25, Corollary 3 and Theorem 1], where $\sigma_{\text{med}}$ is the median of the observed singular value distribution of the weight matrix $A$ and $\mu_{\text{med}}$ is the median of a Marčenko-Pastur probability density function. The median $\mu_{\text{med}}$ is generally unknown, but can be computed as explained in Ref. [25, p.5046]. These results, based on random matrix theory [22], are all rigorous in an asymptotic framework under specific conditions given in Ref. [25]. We define $\text{thr ank}(A)$ has the number of singular values above the optimal singular value threshold $\tau^*(A)$, i.e.,

$$
\text{thr ank}(A) = \# \{ \sigma_i \mid i \in \{1, \ldots, N\} \text{ and } \sigma_i > \tau^*(A) \}, \tag{S41}
$$

where $\#$ is the cardinal of a set.

- In a similar spirit as the optimal threshold, one can consider the optimal shrinkage of singular values [27, 49, 55] to define an effective rank. Let $A = A_\ell + X$ where $A_\ell$ is a (deterministic) matrix of unknown rank $\ell$ and $X$ is some noise random matrix. Shortly, given the singular values of $A$, the scalar function $s : [0, \infty) \rightarrow [0, \infty)$, $\sigma_i \mapsto s(\sigma_i)$ is called a **shrinker** or a **denoiser** of singular values. From Refs. [27, 49, 55], 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 $\text{sh r ank}$ has the rank of the matrix with optimally shrinked singular values, i.e.,

$$
\text{sh r ank}(A) = \# \{ s^*(\sigma_i) \mid i \in \{1, \ldots, N\} \text{ and } s^*(\sigma_i) > 0 \}. \tag{S42}
$$

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. [27].

**Remark S13.**

- 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 [56] (e.g., with sklearn.utils.extmath.randomized_svd in Python).
- In Fig. S2, we observe a slow linear growth of the effective ranks as the number of vertices increases.

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

**Remark S14.**

- There is already a low-rank hypothesis hidden in the definition of the optimal threshold and shrinkage, since it is assumed that the rank is finite in the limit $N \rightarrow \infty$. Yet, as observed experimentally in the paper, the finite rank hypothesis is a **posteriori** justified.
- If the type of noise is unknown, the assumptions of Refs. [25, 27] 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.[27]. Moreover, we correct an error made in Ref.[27] concerning the optimal singular value shrinkage for operator norm loss with the Theorem 3.1 of W. Leeb [55]. 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 [25]. 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 [22, 25, 27, 49, 55] will have a
considerable impact on network science (besides, it already has an impact in neuroscience [57]). Indeed, considering noisy networks is a long-standing challenge in network science (e.g., in sociology [58]) that has been addressed, for instance, with Bayesian inference [59–62]. 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 [27] 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 now 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 derive some intuitive inequalities for the stable rank of graphs and show that the exponential decay of singular values naturally leads to low effective ranks. Second, we highlight the ubiquity of the low-rank hypothesis in random graph theory. Third, we present an original theorem for the rapid decrease of the singular values in the directed soft configuration model. 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. Stable rank of graphs

The stable rank, abbreviated srank, has been defined in Sec. I C as the ratio of the Frobenius norm squared over the spectral norm squared. Below, we take advantage of this simple formula, which is amenable for analytic calculations, to derive many useful inequalities. This will ultimately lead to the conclusions that the stable rank increases linearly or sub-linearly with the number of vertices in large graphs and that exponentially decreasing singular values imply low stable rank.

Lemma S15. Let $A$ be a $N \times N$ matrix. Then the Frobenius norm of $A$ is upper bounded as

$$\|A\|_F \leq N \max_{i,j} |A_{ij}|.$$ (S43)

Moreover, the spectral norm of $A$ is lower bounded as

$$\|A\|_2 \geq \max \left\{ \max_i \|r_i\|, \max_j \|c_j\|, \frac{1}{\sqrt{N}} \|k^{\text{in}}\|, \frac{1}{\sqrt{N}} \|k^{\text{out}}\| \right\},$$ (S44)

where $r_i$ and $c_j$ respectively denote the $i$-th row and $j$-th column of $A$ while $k^{\text{in}} = A1$ and $k^{\text{out}} = A^T 1$. 

Fig. S2: Different effective ranks vs. the number of vertices for 674 networks of diverse origins.
Proof. The first inequality trivially follows from the definition of the Frobenius norm:

\[ \|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2} \leq \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{NA_{\max}} = N A_{\max}, \quad A_{\max} = \max_{i,j} |A_{ij}|. \]  

(S45)

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

\[ \|A\|_2 = \max_{\|x\|=1} \|Ax\|, \]  

(S46)

which implies that \( \|A\|_2 \geq \|Ax\| \) for any \( x \) such that \( \|x\| = 1 \). Choosing \( x = e_j \), the \( j \)-th unit vector, leads to the inequality \( \|A\|_2 \geq \|e_j\| \). But this inequality holds all \( j \), so we can combine all the inequalities and infer that

\[ \|A\|_2 \geq \max_j \|e_j\|. \]  

(S47)

Now, because the spectral norm is invariant under matrix transposition, we also know that \( \|A\|_2 \geq \|A^\top x\| \) for any \( x \) such that \( \|x\| = 1 \). Setting once again \( x = e_i \) for all \( i \), we conclude that

\[ \|A\|_2 \geq \max_i \|r_i\|. \]  

(S48)

Choosing \( x = 1/\sqrt{N} \) in \( \|A\|_2 \geq \|Ax\| \) and in \( \|A\|_2 \geq \|A^\top x\| \) readily provides the two other lower bounds.

Proposition S16. Let \( A \) 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,

\[ \text{srank}(A) \leq \frac{M}{k_{\max}}. \]  

(S49)

Proof. We first note that when \( A \) 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,

\[ \|A\|_F^2 = \sum_{i,j} A_{ij}^2 = \sum_{(i,j)} 1 = M. \]  

(S50)

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

\[ \|A\|_2^2 \geq \max_i \left\{ \max_j A_{ij}^2, \max_i \sum_j A_{ij} \right\} = \max_i \sum_j A_{ij}, \max_i \sum_j A_{ij} \right\} = \max_i \left\{ \max_{i,j} k_i^\text{in}, \max_{j,i} k_j^\text{out} \right\} = k_{\max}. \]  

(S51)

Thus, \( \text{srank}(A) = \|A\|_F^2/\|A\|_2^2 \leq M/k_{\max} \), 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 \( M = O(N^{2-\epsilon}) \) and \( k_{\max} = O(N^{1-\epsilon}) \) for some \( \epsilon > \epsilon > 0 \), then the stable rank is of order \( O(N^{1+\epsilon-\epsilon}) = o(N) \). As shown in next proposition, similar scaling behaviors emerge when considering general square matrices, which are relevant for studying signed weighted directed graphs.

Proposition S17. Let \( p \geq 0 \) and \( \alpha, \beta > 0 \). Let \( A \) be a \( N \times N \) matrix such that

\[ \alpha N^{-p} \leq A_{ij}^2 \leq \beta N^{-p} \]  

(S52)

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

\[ \text{srank}(A) \leq \alpha^{-1} \beta N. \]  

(S53)

More generally, if the maximum number of nonzero elements in a row or in a column of \( A \) is \( \gamma N \), the total number of nonzero elements of \( A \) is \( \delta N^2 \), and all these nonzero elements satisfy inequality \( (S52) \), then

\[ \text{srank}(A) \leq \alpha^{-1} \beta \gamma^{-1} \delta N. \]  

(S54)

Proof. We use Lemma S15 and proceed essentially as for Proposition S16.

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 \( \gamma = \tilde{\gamma} N^{-\epsilon} \) and \( \delta = \tilde{\delta} N^{-\epsilon} \) for some \( \epsilon > \epsilon > 0 \), which implies that once again, \( \text{srank} = O(N^{1+\epsilon-\epsilon}) = o(N) \). A typical sparse matrix has \( \epsilon = 2\epsilon \), leading to a stable rank scaling as \( O(N^{1-\epsilon}) \), which tends to \( O(1) \) when considering the ultra-sparse case \( \epsilon = 1 \).
We have proved so far that the stable rank of (signed weighted directed) graphs having \( N \) vertices tends to increase linearly with \( N \) and that sparsity makes the increase become sub-linear. This means that sparse graphs are characterized by a ratio \( \text{srank}/N \) that goes to zero as \( N \) grows, obviously corresponding to a low effective rank. We are now going to prove that low effective rank also results from exponentially decreasing singular values, a phenomenon that we have observed in many real and synthetic graphs and that is not correlated to sparsity from our numerical experiments (not shown here).

**Lemma S18.** Let \( A \) be a \( N \times N \) matrix whose first singular value satisfy \( \sigma_1 \geq \alpha N^{1-\epsilon} \) for some \( \alpha > 0 \) and \( 0 \leq \epsilon \leq 1 \). Assume that all the singular values of \( A \) are upper-bounded by an exponentially decreasing factor as

\[
\sigma_i \leq \beta N^{1-\epsilon} \gamma^i, \quad i \in \{1, \ldots, N\},
\]

where \( \beta > 0 \) and \( 0 < \gamma < 1 \). Then,

\[
\text{srank}(A) \leq \frac{\beta^2 \gamma^2}{\alpha^2} \frac{1 - \gamma^{2N}}{1 - \gamma^2} < \frac{\beta^2}{\alpha^2} \gamma^2.
\]

**Proof.** First, we know that \( \text{srank}(A) = \sum_{i=1}^{N} (\sigma_i/\sigma_1)^2. \) Second, we use the lower bound on \( \sigma_i \) and the upper bound on \( \sigma_i \) to get the inequality

\[
\text{srank}(A) \leq \frac{\beta^2}{\alpha^2} \sum_{i=1}^{N} \gamma^{2i} = \frac{\beta^2 \gamma^2}{\alpha^2} \sum_{k=0}^{N-1} \gamma^{2k}.
\]

We finally use the formula for the finite geometric series and the trivial inequality \( 1 - \gamma^{2N} < 1 \). \( \square \)

**Remark S19.** In the previous lemma, the singular values scale as \( \sigma_i \sim N^{1-\epsilon} \) when \( N \rightarrow \infty \). This can be easily justified from the fact that each singular value is a solution to an optimization problem of the form

\[
\sigma_i = \max_{\parallel x \parallel=1 \text{orthogonality conditions}} \parallel A x \parallel,
\]

which asymptotically behaves as \( N^{1-\epsilon} \) whenever \( A_{ij} \sim N^{-\epsilon} \). The parameter \( \epsilon \) used in the theorem can thus be interpreted as a sparsity factor, which has no impact on the scaling of the stable rank.

The last lemma states that the asymptotic behavior \( \text{srank}(A) = O(1) \) emerges from the exponential decay of the singular values. In other words, observing exponentially decreasing singular values in a large graph is sufficient to conclude that \( \text{srank}(A)/N = o(1) \), which is negligible compared to \( N \), and thus indicates that the effective rank of the graph is low. In the next subsection, we will show that many models of random graphs enjoy an even stronger asymptotic behavior, namely \( \text{rank}(A) = O(1) \). We will also numerically verify the exponential decay of the singular values for three important random graph models (FIG. S3). Later in subsection II C, we will analytically prove that the singular values of the directed soft configuration model exponentially decrease as in Eq. (S55).

**B. SVD of random graphs**

Random graphs and their eigenvalue spectrum have a long and rich history [48, 63–74], but less attention has been given to their singular value decomposition. The adjacency or the weight matrix of a random graph model can always be written as

\[
A = \langle A \rangle + X,
\]

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

\[
\langle A \rangle = \Phi(L),
\]

where \( \Phi \) 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 \( \phi \) being a real scalar function of a real variable. In the next example, we give some precision about the random graphs of Table I in the Methods.

**Example S20** (Network science). A large class of binary random graphs are described by Bernouilli random matrices, \( X_{ij} \) being equal to either \(-\langle A_{ij} \rangle\) or \( 1 - \langle A_{ij} \rangle\). The expected adjacency matrix for...

- ...the \( G(N, p) \) model [75–77] with self-loops is

\[
\langle A \rangle = L = Np \; \mathbf{1} \mathbf{1}^\top,
\]

where \( \mathbf{1} \) is a row vector of ones.
where $Np\hat{1}\hat{1}^T$ 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 $\hat{1} = (1 \ldots 1)^T/\sqrt{N}$. The model is also called Poisson random graph, Erdős-Rényi model [78], Bernoulli random matrix [79], or spiked Wigner matrix [80].

- The **stochastic block model (SBM)** [81, 82] with $q$ communities (generalization of $G(N,p)$) is

$$\langle A \rangle = L = \sum_{\mu, \nu = 1}^q \sqrt{p_{\mu \nu}} b_\mu b_\nu^T,$$

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

- The **random dot product graph** [83] is

$$\langle A \rangle = XX^T,$$

where $X$ is a $N \times d$ matrix where the rows are the latent positions of each vertex of the graph. The rank of $\langle A \rangle$ is obviously $d$. The model generalizes the SBM, the degree-connected SBM and the mixed-membership SBM as shown in Theorem 15 of Ref. [83].

- The **Chung-Lu model** [84, 85] is

$$\langle A \rangle = L = \frac{||\kappa||^2}{2M} \bar{\kappa} \bar{\kappa}^T,$$

where $\kappa$ is a vector of expected degrees. Note that the annealed approximation, omnipresent in epidemiology [86], is thus a very strong low-rank hypothesis (effective rank 1) for complex networks.

- The **metadegree model** [87] (generalization of the Chung-Lu model) is

$$\langle A \rangle = L = \sum_{\mu, \nu = 1}^r \Delta_{\mu \nu} v_\mu v_\nu^T,$$

where $(v_\mu)_{\mu = 1}^r$ are the $N$-dimensional vectors of metadegree and $\Delta$ is a $r \times r$ nonsingular matrix that contains the “coefficients of mixing” among metadegrees. In Ref. [87, p.2, 2nd column, 2nd paragraph], a low-rank hypothesis is explicitly made as they assume that the rank of $V\Delta V^T$ is much smaller than the size of the system.

- The $S^1$ model of **random geometric networks** [88] is such that

$$\langle A_{ij} \rangle = \phi(L_{ij}) = \frac{1}{1 + L_{ij}^{\beta}},$$

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

$$L_{ij} = \frac{R\theta_{ij}}{\mu \kappa_i \kappa_j},$$

where $\mu$ and $R$ are positive constants, the latter representing the radius of the circle on which the vertices are distributed, $\theta_{ij}$ is the angular distance between the vertices $i$ and $j$ on the circle, and $\kappa_i$ denotes the $i$-th latent degree, yet another positive constant. To estimate the rank of the matrix $L$, it is more convenient to rewrite it using the Hadamard product:

$$L = \frac{R}{\mu} \bar{\kappa} \bar{\kappa}^T \circ \theta,$$

where

$$\bar{\kappa} = (1/\kappa_i)_{i=1}^N, \quad \theta = (\theta_{ij})_{i,j=1}^N.$$

Clearly, $\bar{\kappa} \bar{\kappa}^T$ is a rank-one matrix. Now, according to Ref. [89], the rank of a distance matrix is at most $D + 2$, where $D$ is the dimension of the manifold where the points are embedded. Here $D = 1$, which means that the rank of $\theta$ 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 rank at most 3.

- The **soft directed configuration model** is such that

$$\langle A_{ij} \rangle = \phi(L_{ij}) = \frac{L_{ij}}{1 + L_{ij}}, \quad L_{ij} = \alpha_i \beta_j,$$

for some positive parameters $\alpha_i, \beta_j$. In other words, $L = \alpha \beta^T$ is a rank-one matrix where while $\alpha, \beta$ are positive vectors defined in Section II C.

- The **Barabási-Albert model (BA)** [90], a model of preferential attachment and a particular case of Price’s
model [91], does not possess an explicit formula of the form $\langle A \rangle = \Phi(L)$. However, we observe numerically that as the number of edges $m$ to which a new vertex is attached increases (for $m \geq 2$), the stable rank, the nuclear rank, the elbow, thrank, and shrank tend to decrease while the rank, the erank, and the energy ratio tend to increase.

In Fig. S3, we show the singular values of the SBM, the BA model, and the $S^1$ random geometric model.

![Fig. S3: Rescaled singular values for an instance of three different random graphs with 10^3 vertices (a,b,c) and the related spectral densities $\rho(\sigma)$ for 10^3 instances (d,e,f) with a zoom on the largest singular values (insets). a-d, the directed stochastic block model with self-loops, five blocks with sizes [100, 400, 100, 200, 200], and probabilities of connections inside and outside the blocks [[0.4, 0.1, 0.3, 0.01, 0.001], [0.05, 0.6, 0.2, 0.1, 0.05], [0.2, 0.01, 0.7, 0.2, 0.01], [0.15, 0.05, 0.05, 0.8, 0.01], [0.01, 0.1, 0.05, 0.2, 0.5]], where an element $(\mu, \nu)$ is the probability $p_{\mu\nu}$ of a connection from block $\nu$ to $\mu$. b-e, the Barabási-Albert model with two edges to attach from a new vertex to an existing vertex. c-f, the (directed) $S^1$ random geometric model where $\beta = 2.5$ controls the clustering, the expected degrees are sampled from a truncated Pareto probability density function with minimum expected degree $\kappa_{\min} = 3$, maximum expected degree $\kappa_{\max} = 20$, and shape parameter $s = 1.5$. The zero singular values are not illustrated in the scree plots b and c.]

| Model | rank | nrank | elbow | energy | thrank | shrank | erank | rank |
|-------|------|-------|-------|--------|--------|--------|-------|------|
| SBM   | 3    | 37    | 5     | 374    | 5      | 29     | 707   | 1000 |
| BA    | 38   | 140   | 102   | 372    | 86     | 131    | 636   | 822  |
| $S^1$ | 36   | 137   | 117   | 362    | 97     | 135    | 686   | 993  |

There are also counter-examples of the low-rank hypothesis for networks. The most obvious example is perhaps the Watts-Strogatz model, which had a considerable impact in the field.

**Example S21.** The Watts-Strogatz model [92] is described by the random matrix

$$A = D_k + X$$

where $D_k$ is a band matrix of bandwidth $k$ whose $k$ up- and sub-diagonal entries are equal to 1 while $X$ is a matrix 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 graph models are low-rank”: the model is a sum of a high-rank matrix and a noise matrix. This observation confirms that even though the Watt-Strogatz satisfy two interesting properties, namely small distances (small-world property) and a high-clustering coefficient, it doesn’t enjoy the low-effective-rank property that we observe in real networks.
There are also clear examples of the low-rank hypothesis in physics, machine learning and neuroscience that are worth mentioning.

**Example S22** (Random matrix theory and spin glasses). The typical random matrix ensembles used in physics (e.g., Gaussian Orthogonal Ensemble [28]) are matrix models involving normally (Gaussian) distributed random variables and such that \( \langle A \rangle = 0 \), 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 [93] 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** [94] or **spiked random matrices** [95], satisfy Eq. (S58) with \( \langle A \rangle \) of rank \( r = o(N) \), ensuring their compatibility with the low-rank hypothesis in the limit \( N \to \infty \).

**Example S23** (Machine learning and neuroscience). (1) In the **Hopfield network** [96], 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 \( \mathbf{v} \in \{0,1\}^N \). Starting with an initial random symmetric weight matrix of mean zero, \( T \), the training consists in mapping \( T \mapsto T + \sum_{s=1}^n \mathbf{v}_s \mathbf{v}_s^\top \), resulting in a final weight matrix of effective rank \( \leq n \) that trivially complies with the low-rank hypothesis. (2) In an **echo-state network** [97], 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). (3) Training **shallow undercomplete autoencoders** is essentially a low-rank approximation problem [20]. The architecture formed by encoding/decoding weight matrices and a hidden layer thus form a low-rank model in itself. (4) Another example is the **Gaussian mixture low-rank network** [98–101] whose weight matrix is defined as

\[
A = \frac{1}{N} \sum_{r=1}^R \mathbf{m}_r \mathbf{n}_r^\top + X,
\]

where \( X \) is a zero-mean Gaussian random matrix while \( \mathbf{m}_r \) and \( \mathbf{n}_r \) respectively denote the \( r \)-th left singular vector of \( \langle A \rangle \) and its right singular vector multiplied by the \( r \)-th singular value. The rank of \( \langle A \rangle \) is thus equal to \( R \). In Ref. [100], 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 \( R \ll N \).” It has also been observed experimentally that trained models have a low effective rank (the ones from NWS [102] in this paper and other references, such as Ref. [103]).

Despite the clear dependence of \( \langle A \rangle \) over a low-rank matrix \( L \), it is not always clear whether \( \langle A \rangle \) has an effective low-rank. This is the case of **soft configuration models** (see Example S20) 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, characteristic of the low-rank hypothesis.

**C. Exponential decrease of singular values in the directed soft configuration model**

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 [104]. A lot of random graphs are defined from a maximally entropic model and although there is a large literature on the subject [105–108], we provide, for the sake of completeness, some important results and comments. We will later use them to demonstrate Theorem S34 on the exponential decrease of singular values in the directed soft configuration model, a maximally entropic random graph model.

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 [109, 110]. 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 [111, 186 and 187].

**Theorem S24** (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 \( x \in E \) if \( x \in U \) and \( g_1(x) = \cdots = g_r(x) = 0 \).

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

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

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

\[
\nabla f(x^*) + \sum_{i=1}^{r} \lambda_i^* \nabla g_i(x^*) = 0
\]

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

**Proof.** The proof is long and often based on the local inversion theorem. See Carathéodory multiplicative rule in [112] or [113, Theorem 20.3].

**Remark S25.** On the one hand, the theorem is valid for a minimum or a maximum. This is an advantage that can turn out to be a inconvenient if we do not verify the nature of the point \( x^* \). On the other hand, Eq. (S72) is only a necessary condition and it is not sufficient in general. We could, for example, find a solution of Eq. (S72) 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 \( D = [0,1]^N \). Finally, the gradients of the constraints must be linearly independent; otherwise \( \lambda_0 \) can be 0 and that does not help to find the extremum.

The Lagrange multiplier method begins by solving Eq. (S72) by expressing all \( x_i^* \) in terms of the multipliers \( \lambda_i \). Then, the multipliers are written in terms of the known variables by solving the set of constraints \( g_1(x) = \cdots = g_n(x) = 0 \), which is generally the most difficult step. We finally verify that the solution \( x^* \) 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 S26.** Let \( A \) be a \( N \times N \) random adjacency matrix with support \( \Omega_A \) that satisfies the soft equality constraints

\[
E[ h_\mu(A)] = h_\mu(\mu^*), \quad \mu \in \{1, \ldots, \ell\},
\]

where \( E \) is the expected value on \( \Omega_A \), \( \mu^* \) is some \( N \times N \) non-random adjacency matrix, and each \( h_\mu: \{0,1\}^{N^2} \rightarrow \mathbb{R}^\ell \) is continuously differentiable. Then, the probability mass function \( P \) that maximizes the entropy of \( A \) under the equality constraints (S73) must be of the form

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

where \( Z: \mathbb{R}^\ell \rightarrow \mathbb{R} \) is the partition function and \( \lambda_\mu \neq 0 \) for all \( \mu \).

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. [105].

**Corollary S27.** Let \( k^{\text{in}} \) and \( k^{\text{out}} \) 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:

\[
E[A1] = k^{\text{in}} \quad \text{and} \quad E[A^T1] = k^{\text{out}}.
\]

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

\[
P(a) = \prod_{i,j=1}^N p_{ij}^{\alpha_{ij}} (1 - p_{ij})^{1 - \alpha_{ij}}, \quad p_{ij} = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j},
\]

where \( \alpha_i \) and \( \beta_j \) are positive numbers \( \forall i, j \in \{1, \ldots, N\} \).

**Remark S28.** The 2N scalars \( (\alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_N) \) are such that \( \alpha_i = e^{\lambda_i} \) and \( \beta_j = e^{\lambda_{N+i}} \) for all \( i, j \in \{1, \ldots, N\} \) where \( \lambda_1, \ldots, \lambda_N \) are the Lagrange multipliers related to the in-degree constraints, while \( \lambda_{N+1}, \ldots, \lambda_{2N} \) 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.
Corollary S29. Let $A$ be the random matrix described in the previous corollary. Then, for all $i, j \in \{1, \ldots, N\}$,
\[
\langle A_{ij} \rangle = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j} \tag{S77}
\]
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 $\ell_1, \ell_2, \ldots$ or $N, m_1, m_2, m_3, \ldots$

Lemma S30. Let $A$ be a random matrix satisfying Eq. (S77). Let
\[
\alpha = (\alpha_1, \ldots, \alpha_N)^T, \quad \beta = (\beta_1, \ldots, \beta_N)^T.
\]
(1) If $0 < \alpha_i \beta_j < 1$ for all $i, j \in \{1, \ldots, N\}$, then
\[
\langle A \rangle = \sum_{k=1}^{\infty} L_k, \tag{S78}
\]
where $L_k$ denotes a rank-one $N \times N$ matrix whose only nonzero singular value is
\[
\ell_k = \sqrt{\sum_{i,j=1}^{N} (\alpha_i \beta_j)^{2k}}. \tag{S79}
\]
(2) If $\alpha_i \beta_j > 1$ for all $i, j \in \{1, \ldots, N\}$, then
\[
\langle A \rangle = N 11^T + \sum_{k=1}^{\infty} M_k, \tag{S80}
\]
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 $i, j \in \{1, \ldots, N\}$, we can use the geometric series and get the following convergent series:
\[
\langle A \rangle = \alpha \beta^T - (\alpha \circ \alpha) (\beta \circ \beta)^T + (\alpha \circ \alpha \circ \alpha) (\beta \circ \beta \circ \beta)^T - + \ldots
\]
Setting
\[
L_k = (-1)^{k+1} \underbrace{(\alpha \circ \cdots \circ \alpha)}_{k \text{ times}} \underbrace{(\beta \circ \cdots \circ \beta)^T}_{k \text{ times}},
\]
we get Eq. (S78). We see that each matrix $L_k$ is factorized as $uv^T$, so we conclude that the rank of each element of the series is one. Moreover, the SVD for such a matrix is simply $uu^T = \rho \hat{u} \hat{v}^T$, where $\rho = \|u\|\|v\|$, $\hat{u} = u/\|u\|$, $\hat{v} = v/\|v\|$. Hence,
\[
L_k = \ell_k \hat{\alpha}_k \hat{\beta}_k^T, \tag{S80}
\]
where
\[
\ell_k = \| \underbrace{\alpha \circ \cdots \circ \alpha}_{k \text{ times}} \| \| \underbrace{\beta \circ \cdots \circ \beta}_{k \text{ times}} \|,
\quad \hat{\alpha}_k = (-1)^{k+1} \underbrace{\alpha \circ \cdots \circ \alpha}_{k \text{ times}} \| \| \underbrace{\beta \circ \cdots \circ \beta}_{k \text{ times}} \|,
\quad \hat{\beta}_k = \underbrace{\beta \circ \cdots \circ \beta}_{k \text{ times}} \| \| \underbrace{\alpha \circ \cdots \circ \alpha}_{k \text{ times}} \|.
\]
Simple calculations lead to Eq. (S79), 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 $\ell_k$ and $m_k$ in Lemma S30 are ordered and bounded as stated in the next lemma.
Lemma S31. Let $\ell_k$ and $m_k$ be the coefficients defined in Lemma S30.

(1) If $0 < \alpha_i \beta_j < 1$ for all $i \in \{1, \ldots, N\}$, then
\[ \ell_{k+1} < \ell_k , \quad \forall \ k \in \mathbb{Z}_+ . \] (S81)

If $\alpha_i \beta_j < \gamma$ for all $i, j \in \{1, \ldots, N\}$, for some $\gamma \in (0, 1)$, then
\[ \ell_k < N \gamma^k , \quad \forall \ k \in \mathbb{Z}_+ . \] (S82)

(2) If $\alpha_i \beta_j > 1$ for all $i, j \in \{1, \ldots, N\}$, then
\[ m_{k+1} < m_k , \quad \forall \ k \in \mathbb{Z}_+ . \] (S83)

If $\alpha_i \beta_j > \omega$ for all $i \in \{1, \ldots, N\}$, for some $\omega > 1$, then
\[ m_k < N \omega^{-k} , \quad \forall \ k \in \mathbb{Z}_+ . \] (S84)

Proof. For the first case, for all $k \in \mathbb{Z}_+$ and from Eq. (S79),
\[ \ell_{k+1} = \sqrt{\sum_{i,j=1}^N (\alpha_i \beta_j)^{2(k+1)}} = \sqrt{\sum_{i,j=1}^N (\alpha_i \beta_j)^{2k}(\alpha_i \beta_j)^2} < \sqrt{\sum_{i,j=1}^N (\alpha_i \beta_j)^{2k}} = \ell_k , \]
where we have used $(\alpha_i \beta_j)^2 < 1$ since $\alpha_i \beta_j < 1$ for all $i, j \in \{1, \ldots, N\}$. Moreover, if $\alpha_i \beta_j < \gamma$ for all $i, j \in \{1, \ldots, N\}$, then
\[ \ell_k = \sqrt{\sum_{i,j=1}^N (\alpha_i \beta_j)^{2k}} < \sqrt{\sum_{i,j=1}^N \gamma^{2k}} = N \gamma^k . \] (S85)

Case (2) is proven similarly. \( \square \)

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

Lemma S32. Let $A$ be a random matrix satisfying Eq. (S77). Let $\gamma$ and $\omega$ be two positive constants. Then,
\[ \alpha_i \beta_j < \gamma < 1 \iff \langle A_{ij} \rangle < \frac{\gamma}{1+\gamma} < \frac{1}{2} \quad \text{and} \quad \alpha_i \beta_j > \omega > 1 \iff \langle A_{ij} \rangle > \frac{\omega}{1+\omega} > \frac{1}{2} . \] (S86)

Proof. The first part of both equivalences is obtained with basic inequality manipulations:
\[ \alpha_i \beta_j < \gamma \iff \frac{1}{\alpha_i \beta_j} > \frac{1}{\gamma} \iff \frac{1}{1+\frac{1}{\alpha_i \beta_j}} < \frac{1}{1+\gamma} \iff \langle A_{ij} \rangle < \frac{\gamma}{1+\gamma} , \]
\[ \alpha_i \beta_j > \omega \iff \frac{1}{\alpha_i \beta_j} < \frac{1}{\omega} \iff \frac{1}{1+\frac{1}{\alpha_i \beta_j}} > \frac{1}{1+\frac{1}{\omega}} \iff \langle A_{ij} \rangle > \frac{\omega}{1+\omega} . \]

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

Remark S33. The inequalities in the last lemma imply that for all $i \in \{1, \ldots, N\}$, the expected degrees $k_i^{\text{in}}$ and $k_i^{\text{out}}$ 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 $k_i^{\text{in}}$ and $k_i^{\text{out}}$ do not necessarily imply that the inequalities in the last lemma are satisfied. We are now ready to present the 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 S34. Let $(A)$ be the $N \times N$ matrix defined in Eq. (S77) and whose singular values are $\sigma_1 \geq \ldots \geq \sigma_N$. Let $\ell_k$ and $m_k$ be the coefficients defined in Lemma S30.

(1) If $0 < \langle A_{ij} \rangle < 1/2$ for all $i, j \in \{1, \ldots, N\}$, then the singular values are upper bounded convergent series as
\[ \sigma_i \leq \sum_{k=1}^\infty \ell_k , \quad \forall \ i \in \{1, ..., N\} . \] (S87)

If $\langle A_{ij} \rangle < \gamma/(1+\gamma)$ for all $i, j \in \{1, \ldots, N\}$ and for some $\gamma \in (0, 1)$, then the singular values are upper bounded by exponentially decreasing terms, i.e.,
\[ \sigma_i \leq \frac{N \gamma^i}{1-\gamma} , \quad \forall \ i \in \{1, ..., N\} . \] (S88)
(2) If $1/2 < \langle A_{ij} \rangle < 1$ for all $i \in \{1, \ldots, N\}$, then the singular values are upper bounded by convergent series as

$$
\sigma_i \leq N \delta_{i1} + \sum_{k=1}^{\infty} m_k, \quad \forall \ i \in \{1, \ldots, N\},
$$

(S89)

where $\delta_{i1}$ is a Kronecker delta. If $\langle A_{ij} \rangle > \omega/(1 + \omega)$ for all $i, j \in \{1, \ldots, N\}$ and for some $\omega > 1$, then the singular values are upper bounded by exponentially decreasing terms, i.e.,

$$
\sigma_i \leq N \delta_{i1} + \frac{N \omega^{1-i}}{\omega - 1}, \quad \forall \ i \in \{1, \ldots, N\}.
$$

(S90)

Proof.

(1) First of all, $0 < \langle A_{ij} \rangle < 1/2$ if and only if $0 < \alpha_i \beta_j < 1$ for all $i, j \in \{1, \ldots, N\}$ from Corollary S29 and Lemma S32. Lemma S30 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:

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

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

Next, recall from Theorem S8 that the Weyl inequalities for $N \times N$ matrices $B$ and $C$ are

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

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

$$
\sigma_1(B + C) \leq \sigma_1(B) + \sigma_1(C). 
$$

(S91)

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

$$
\sigma_1 \left( \sum_{k=1}^{n} L_k \right) \leq \sum_{k=1}^{n} \sigma_1(L_k).
$$

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

$$
\sum_{k=1}^{n} \sigma_1(L_k) \leq \sum_{k=1}^{n+1} \sigma_1(L_k) \leq \cdots \leq \sum_{k=1}^{\infty} \sigma_1(L_k).
$$

Consequently, for all $1 \leq i \leq n - 1 < \infty$,

$$
\sigma_1 \left( \sum_{k=1}^{n} L_k \right) \leq \sum_{k=1}^{\infty} \sigma_1(L_k).
$$

(S92)

The continuity of $\sigma_1 : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}$, which is obvious since $\sigma_1$ is a norm, and the convergence of $\sum_{k=1}^{\infty} L_k$ allow us to take the limit $n \rightarrow \infty$ on the left-hand side of the previous inequality and conclude that

$$
\sigma_1 \left( \sum_{k=1}^{\infty} L_k \right) \leq \sum_{k=1}^{\infty} \ell_k, \quad \forall \ i \in \mathbb{Z}_+, 
$$

(S93)

where we used the notation $\ell_k = \sigma_1(L_k)$ introduced in Lemma S30.

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

$$
\sigma_1 \left( \sum_{k=1}^{i} L_k \right) = \sigma_1 \left( \sum_{k=1}^{i-1} L_k + \sum_{k=i}^{\infty} L_k \right) \leq \sigma_1 \left( \sum_{k=1}^{i-1} L_k \right) + \sigma_1 \left( \sum_{k=i}^{\infty} L_k \right),
$$

(S94)

which is valid for all $1 \leq i \leq N$. 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, $\sigma_i \left( \sum_{k=1}^{i-1} L_k \right) = 0$, so that

$$
\sigma_1 \left( \sum_{k=1}^{\infty} L_k \right) \leq \sigma_1 \left( \sum_{k=1}^{\infty} L_k \right), \quad \forall \ i \in \{1, \ldots, N\}.
$$

(S95)

Combining inequalities (S92) and (S94) leads to the desired result:

$$
\sigma_1 \left( \sum_{k=1}^{\infty} L_k \right) \leq \lim_{n \rightarrow \infty} \ell_k, \quad \forall \ i \in \{1, \ldots, N\}.
$$

(S95)
We stress that both sides of the inequality are finite. Indeed, on the one hand, the series $\sum_{k=1}^{\infty} L_k$ is convergent. On the other hand, $\ell_k > 0$ and by Lemma S31, the inequality $\ell_{k+1} < \ell_k$ holds for all $k$. Then,

$$L = \lim_{k \to \infty} \left| \frac{\ell_{k+1}}{\ell_k} \right| = \lim_{k \to \infty} \frac{\ell_{k+1}}{\ell_k} < 1$$  \hspace{1cm} (S96)

and d’Alembert’s ratio test ensures that the infinite series $\sum_{k=1}^{\infty} \ell_k$ is convergent for all $i$. To get a more explicit inequality, we recall from Lemmas S31–S32 that if $\langle A_{ij} \rangle < \gamma / (1 + \gamma)$, then $\ell_k < N \gamma^k$. This in turn implies that $\sum_{k=1}^{\infty} \ell_k < N \gamma^k / (1 - \gamma)$. Consequently,

$$\sigma_i \left( \sum_{k=1}^{\infty} L_k \right) < \frac{N \gamma^i}{1 - \gamma}, \quad \forall i \in \mathbb{Z}_+$$

(2) Similarly to the first case, Corollary S29, Lemmas S30–S32, and Weyl inequalities imply that

$$\sigma_i(\langle A \rangle) = \sigma_i \left( \sum_{k=1}^{\infty} M_k \right) \leq N \delta_{i1} + m_i + \sigma_1 \left( \sum_{k=i+1}^{\infty} M_k \right).$$  \hspace{1cm} (S97)

Proceeding as for inequality (S92) then leads to the inequality

$$\sigma_i \left( \sum_{k=1}^{\infty} M_k \right) \leq N \delta_{i1} + \sum_{k=i}^{\infty} m_k,$$  \hspace{1cm} (S98)

where $m_k = \sigma_1(M_k)$. The series $\sum_{k=1}^{\infty} M_k$ is convergent by Lemma S31 while $\sum_{k=i}^{\infty} m_k$ converges by virtue of d’Alembert ratio test. Additionally, if $\langle A_{ij} \rangle > \omega / (1 + \omega)$, then Lemma S32 states that $\alpha_i \beta_j > \omega$. The latter condition and Lemma S31 lead to

$$\sigma_i \left( \sum_{k=1}^{\infty} M_k \right) \leq N \delta_{i1} + N \sum_{k=i}^{\infty} \omega^{-k}.$$  \hspace{1cm} (S99)

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

The bounds in the last theorem theoretically validate the low-rank hypothesis in the expected adjacency matrix of the soft directed configuration model. As shown in Fig. S4, we observe an exponential decrease of the upper bound on the singular values. Even if the upper bound is not necessarily tight for all $\langle A \rangle$, especially for values of $\gamma$ near 1, the observed decrease is clear.

![Fig. S4: Exponential decrease of the singular values of $\langle A \rangle$ in the soft directed configuration model with $10^3$ vertices [circle markers] and its theoretical upper bound [solid lines] based on Weyl inequalities in log-log for different values of $\gamma$ and $\omega$ [Theorem S34]. The dashed lines between singular values are shown for the sake of visualization.](image)

**D. Directed network centrality measures**

The SVD of directed networks is related to two vertex centrality measures: the authority centrality (dominant left singular vector $u_1$) and the hub centrality (dominant right singular vector $v_1$) [114, 115], 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 S39 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 [68, Theorem 38] doesn’t apply).

\[ W = U V^T \]

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. [116]).

E. Adaptative 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 [117]. The effective rank of a complex network should thus
be expected to change according to time. We performed a preliminary investigation of this phenomena by extracting
the effective rank of the \textit{C. elegans} connectome at different stages of its maturation [118] 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.

![Graph showing singular values over developmental stages](image)

Fig. S6: Singular values of the matrices describing the connectivity of the \textit{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. [103], 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. Towards dynamical systems

The applications of SVD for dynamical systems is very broad, especially in engineering and linear control systems [36]. SVD is also generalized for nonlinear operators [119] and it is even possible to perform a quasi-optimal low-rank approximation for matrix dynamics with time-evolving matrices $U, \Sigma, V^T$ [120], which could have interesting applications in the study of temporal networks [121]. 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. [122, 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 [123, p.145][122]). In the following section, we give details about the dimension reduction of complex systems and especially, in Section 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 [124] or getting suitable observable to assess the controllability of the system [125]. 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 [126], coarse graining [127–129], reduced-order model [4], model reduction [130], lumping [131, 132] [133, Section 2.4], compression [134], pruning [135], dominance analysis [136], variable or state aggregation [137], 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;
- $\phi : [0, \infty) \times \mathbb{R}^N \to \mathbb{R}^N$ 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,

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

Consider the following notation for the reduced dynamical system:
- $R : \mathbb{R}^N \to \mathbb{R}^n$ with $n < N$ is called the reduction function or a vectorial observable;
- $R = (R_1, \ldots, R_n)$ where $R_\mu : \mathbb{R}^N \to \mathbb{R}$ is the $\mu$-th observable;
- $X = R(x) \in \mathbb{R}^n$ is a reduced state;
- $\Phi : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ 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. [132], but avoiding the subtleties in the characteristics of the reduction function $R$.

**Definition S35.** The function $R : \mathbb{R}^N \mapsto \mathbb{R}^n$ induces an *exact dimension reduction* of the dynamics

\[ \dot{x} = f(x) \]  

if there exists a vector field $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that for all solutions $x(t)$ of Eq. (S100), the reduced trajectory

\[ X = R \circ x : [0, \infty) \rightarrow \mathbb{R}^n \]  

obeys the differential equation

\[ \dot{X} = F(X). \]  

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 S36.** The following statements are equivalent:

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

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

holds where $\mathcal{U}$ is the Koopman operator generator and $J_R$ is the Jacobian matrix of $R$;
3. the complete flow $\phi_t$ and the reduced flows $\Phi_t$ commutes with $R$ such that

\[ R \circ \phi_t = \Phi_t \circ R. \]  

**Proof.**

(1. $\Leftrightarrow$ 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

\[ \dot{X} = \frac{d(R \circ x)}{dt} = \mathcal{U}[R] \circ x = J_R \circ f \circ x = F \circ R \circ x = F \circ X, \]  

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

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

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

\[ \begin{array}{ccc}
\mathbb{R}^N & \xrightarrow{\Phi_t} & \mathbb{R}^N \\
\downarrow{R} & & \downarrow{R} \\
\mathbb{R}^n & \xrightarrow{\Phi_t \circ R} & \mathbb{R}^n
\end{array} \]  

In Section III D [Definition S38], 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.

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 [122] (or lumping matrix [131, 138]). Then, $J_R = M$ and condition (S103) for closure states that for an exact dimension reduction, the
Fig. S7: Schematization of dimension reduction associated with the reduction matrix $M$, the corresponding projector $P$, and the induced vector subspaces.

complete and reduced vector fields must commute with $M$:

$$\begin{align*}
R^N & \xrightarrow{f} R^N \\
\downarrow M & \downarrow M \circ f \\
R^n & \xrightarrow{F \circ M} R^n \\
\downarrow F & \downarrow F \\
R^n & \xrightarrow{F} R^n
\end{align*}$$

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. [139, p.190 and p.191].

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

$$P = M^+ M,$$

(S109)

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 $W$, the condition $M \circ f = F \circ M$ takes the form of the compatibility equation [122]

$$MW = WM$$

(S110)

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

**Theorem S37** (Ref. [122]). Let $M$ and $W$ be respectively of size $n \times N$ and $N \times N$ with $n < N$. Then, the compatibility equation $WM = MW$ has a solution for $W$ if and only if

$$MWM^+ M = MW$$

(S111)

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

$$W = MWM^+ + Y - YMM^+,$$

(S112)
where $Y$ is an arbitrary $n \times n$ matrix. If $\text{rank } M = n$, then there is at most one solution to the compatibility equation, i.e.,

$$W = MWM^+.$$ (S113)

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

**Proof.** From Penrose 1955 [40, Theorem 2], a necessary and sufficient condition for the equation $AXB = C$ to have a solution is $AA^+ + CB^+ = C$, in which case the general solution is $X = A^+CB^+ + Y(A^+ + AYBB^+)$, where $Y$ is arbitrary. Set $A = I$, $X = W$, $B = M$, $C = MW$ in the latter equations and the result in Eq. (S112) follows. If $\text{rank } M = n$, then the $n$ rows of $M$ are linearly independent. This implies that $MM^+ = I$ and

$$W = MWM^+ + Y - Y = MWM^+,$$

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 [140], that

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

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

As it will be discussed in another paper, the first preliminary results on these compatibility equations seems to go back to 1969 in chemistry [131] and for a fixed $W$, the compatibility equations are homogeneous Sylvester equations (1884) [141].

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 S37.

### 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 [142] 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:

![Diagram](image)

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. 2 for an illustration of $E(x)$].

**Definition S38.** 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.,

\[
\varepsilon(x) = \frac{1}{\sqrt{N}} \|f(x) - M^+F(Mx)\|; \tag{S115}
\]

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

\[
\mathcal{E}(x) = \frac{1}{\sqrt{n}} \|Mf(x) - F(Mx)\|, \tag{S116}
\]

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

\[
\varepsilon'(x) = \frac{1}{\sqrt{N}} \|f(Px) - M^+F(Mx)\| \tag{S117}
\]

and

\[
\mathcal{E}'(x) = \frac{1}{\sqrt{n}} \|Mf(Px) - F(Mx)\| \tag{S118}
\]

in Diagram S114. 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 S35, Theorem S36, and Diagram S108], 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 S39.** 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

\[
\dot{X} = Mf(M^+X) \tag{S119}
\]

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

\[
F^*(X) = \arg \min_{F(X) \in \mathbb{R}^n} \|f(Px) - M^+F(X)\| = Mf(M^+X). \tag{S120}
\]

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, the least-squares theorem (particular case of Penrose [140]) implies that

\[
\arg \min_{u \in \mathbb{R}^\ell} \|v - Au\| = A^+v. \tag{S121}
\]

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 \( \|Mf(Px) - F(X)\| = 0 \). \( \square \)

**Remark S40.**

- 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 \( MW \) \( M^+ \) for the compatibility equation \( MW = WM \) in Theorem S37, 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 \( F = W \).
- 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 \( f : \mathbb{R}^N \rightarrow \mathbb{R}^N \) is a \( C^1(\mathbb{R}^N) \) vector field, the vector fields \( f \) and \( F = M \circ f \circ M^+ = M \circ f \circ M^{-1} \) are \( C^1 \)-conjugate on \( \mathbb{R}^N \) [139, p.191], which is straightforward to observe from the form of \( F \) itself where \( M \) is the \( C^1 \)-diffeomorphism.
- To the authors’ knowledge, even if the vector field in Eq. (S120) is known at least since 1989 [143], 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) [143–146] and especially, on a similar-looking technique for time series which is also loosely [147, 148] called Galerkin projection or Petrov-Galerkin method [4, 36, 145, 149]. In our paper, 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 [150–153] [154, Chap. 11].

**Example S41** (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 \( E_c \) (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 [150] describes the time evolution of the firing rates as

\[
\dot{E}_e = -d_e E_e + (1 - a E_e) S[b(\sum_{e' = 1}^{N_E} W_{ee'} E_{e'} + \sum_{i' = N_E + 1}^{N} W_{ei'} I_{i'} - c_e)] \quad (S122)
\]

\[
\dot{I}_i = -d_i I_i + (1 - a I_i) S[b(\sum_{e'' = 1}^{N_E} W_{ie''} E_{e''} + \sum_{i'' = N_E + 1}^{N} W_{ii''} I_{i''} - c_i)], \quad (S123)
\]

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

\[
S[b(z - c)] = \frac{1}{1 + e^{-(z-c)}} \quad (S124)
\]

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

\[
(x_1,...,x_N) := (E_1,...,E_{N_E},I_{N_E+1},...,I_N)^\top, \quad (S125)
\]

we get a concise form of the model [153, Eq. (11)]:

\[
\dot{x}_j = -d_j x_j + (1 - a x_j) S[b(\gamma y_j - c)], \quad \forall j \in \{1,...,N\}, \quad (S126)
\]

where \( y_j = \sum_{k=1}^{N} W_{jk} x_k \) and we have set \( W \to \gamma W \) to have a coupling constant \( \gamma \) 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 S39, we directly obtain the optimal reduced dynamics

\[
\dot{X}_\mu = \sum_{\nu = 1}^{n} D_{\mu \nu} X_\nu + \sum_{j=1}^{N} M_{\mu j}(1 - a \sum_{\nu = 1}^{n} M_{\nu j}^+ X_\nu) S[b(\gamma \sum_{\nu = 1}^{n} \gamma_{j \nu} X_\nu - c)], \quad (S127)
\]

where \( D_{\mu \nu} = -\sum_{j=1}^{N} M_{\mu j} d_j M_{j \nu}^+ \) and \( \gamma_{j \nu} = \sum_{k=1}^{N} W_{jk} M_{k \nu}^+ \).

Under the form \( Mf(M^+X) \) or, elements by elements, \( \sum_{j=1}^{N} M_{\mu j} f_j(\sum_{\nu = 1}^{n} M_{\nu j}^+ X_\nu) \), 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 [155–159] 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. 2 in the paper]. We demonstrate that the structure that emerges from the dimension reduction in Theorem S39 generally yields higher-order interactions between the observables. For that, we first introduce some assumptions.

**Assumptions S42.**

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

\[
\dot{x}_i = h_i(x_i, y_i), \quad i \in \{1,...,N\}, \quad (S128)
\]

where, for all \( i, x_i : t \to \mathbb{R}^N, y_i = \sum_{j=1}^{N} W_{ij} x_j \), 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 S39]

\[
\dot{X}_\mu = \sum_{i=1}^{N} M_{\mu i} h_i(\tilde{x}_i, \tilde{y}_i), \quad \mu \in \{1,...,n\}, \quad (S129)
\]

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

Condition (1) of Assumptions S42 might look restrictive because of the dependence over the linear function \( x \to Wx \). 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 would even possible to classify the dynamics on networks of the next examples).

**Example S43** (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 \( \gamma \) 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) [133, 160, 161]. 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 [86, 133, 161]. By neglecting the dynamical correlations between the states of the neighbors [162, Sec. 2.3.1], the quench mean-field (QMF) approach [86] yields the deterministic system of equations

\[
\dot{x}_i = -d_i x_i + \gamma (1 - x_i) y_i, \quad i \in \{1, \ldots, N\},
\]

(S130)
called the QMF SIS model, where \( x_i \) is the probability for the vertex \( i \) to be infected. In third figure of the paper, we use the latter dynamics as a simple introductory example. More generally, quench mean-field approximations of many binary stochastic processes, such as the SIS dynamics above, the Cowan dynamics [153, 163], and the Glauber dynamics [164, 165] have the general form

\[
\dot{x}_i = (1 - x_i) \alpha (k_i - y_i) + x_i \beta (k_i - y_i), \quad i \in \{1, \ldots, N\},
\]

(S131)
where \( x_i \) is the probability for vertex \( i \) to be active, \( k_i = \sum_{j=1}^N W_{ij} \) is the in-degree of vertex \( i \), \( \alpha \) (resp. \( \beta \)) is some analytic activation (resp. deactivation) probability function \( \mathbb{R} \to \mathbb{R} \to [0, 1] \).

**Example S44** (Neuroscience). The Wilson-Cowan dynamics in Example S41 satisfies condition (1) of Assumption S42. Another popular model of neuronal activity, the threshold-linear model [166, 167], is defined by the equations

\[
\dot{x}_i = -x_i + \sum_{j=1}^N W_{ij} x_j + b_i, \quad i \in \{1, \ldots, N\},
\]

(S132)
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 \( y \mapsto \ln(1 + e^{\gamma y})/k \) for some \( k > 0 \). 

**Example S45** (Population dynamics). Population dynamics are widely used in science from ecology [168] and game theory [155] to chemistry (e.g., kinetic equations) [169] and physics (e.g., lasers) [170]. The generalized Lotka-Volterra model [171, 172] is a very typical population dynamics with the form

\[
\dot{x}_i = -dx_i + \gamma x_i y_i.
\]

(S133)
Refined models such as

\[
\dot{x}_i = -dx_i - s x_i^2 + \gamma \frac{x_i y_i}{\alpha + y_i}
\]

(S134)
in Ref. [168] or the microbial population dynamics [173]

\[
\dot{x}_i = a - d x_i + b x_i^3 - c x_i^3 + \gamma x_i y_i.
\]

(S135)
have also been used to incorporate more realistic effects, like the Allee effect in which a population exhibits negative growth for low abundances [174, 175]. In the latter dynamics, which is used in the paper, the correspondences with the parameters of Ref. [173] 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 Table III, we consider that the parameter \( d \) can vary for each vertex only for the sake of coherence with the other dynamics.

**Example S46** (Oscillators). The Kuramoto-Sakaguchi dynamics [176, 177] is a canonical model for a large class of oscillatory systems [123, 178] and finds many applications, e.g., for Josephson junctions [179], nanoelectromechanical oscillators [180], and neuroscience [181]. The dynamics of the phase oscillators with a phase lag \( \alpha \) is such that

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

(S136)
where \( \theta_j(t) \) is the position of the \( j \)-th oscillator at time \( t \), \( \omega_j \) is the \( j \)-th natural frequency, and \( \gamma \) is the coupling constant. By setting \( z_j = e^{i \theta_j} \) [122], the Kuramoto-Sakaguchi model becomes

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

(S137)
where \( y_j = \sum_{k=1}^N W_{jk} z_k \) and \( \bar{y}_j \) denotes complex conjugation. Note that the Winfree model [182] and the theta model [183] on networks [122] also satisfy the condition (1) of Assumption S42.
Example S47 (Machine learning). The universal approximation theorem of Funahashi and Nakamura [184, 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 [184, 185]

$$\dot{x}_i = -\frac{1}{\tau_i} x_i + \sum_{j=1}^{N} W_{ij} S(x_j) + I_i,$$  \hspace{1cm} (S138)

where $x_i$ is the trajectory of the $i$-th neuron, $\tau_i$ is the time-scale of neuron $i$, $S$ 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 (S138) is also called or similar, up to some variations in its form, Cohen-Grossberg model [186, 187], Hopfield model [188], activation dynamics [189], continuous rate RNN [190, 191], or reservoir computers. This recurrent neural network does not directly have the form to satisfy the condition (1) of Assumption S42, but from Ref. [192], we know that there is a class of (continuous-time) recurrent neural networks with the form

$$\dot{x}_i = -d_i x_i + \tanh(\gamma y_i + c_i),$$  \hspace{1cm} (S139)

d_1, \ldots, 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 S39.

Proposition S48. If the conditions of Assumptions S42 hold, the reduced dynamics can be expressed in terms of higher-order interactions between the observables as

$$\dot{X}_\mu = C_\mu + \sum_{d_x = 1}^{\infty} \sum_{d_y = 1}^{\infty} D_{\mu\alpha}^{(d_x+1)} X_\alpha + \sum_{d_y = 1}^{\infty} W_{\mu\beta}^{(d_y+1)} X_\beta + \sum_{d_x, d_y = 1}^{\infty} T_{\mu\alpha\beta}^{(d_x+d_y+1)} X_\alpha X_\beta,$$  \hspace{1cm} (S140)

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

$$D_{\mu\alpha}^{(d_x+1)} = \sum_{i=1}^{N} c_{i,0,\alpha} M_{\mu i_0} M_{i_0 a_1}^+ \cdots M_{i_{d_x} a_{d_x}}^+,$$  \hspace{1cm} (S141)

$$W_{\mu\beta}^{(d_y+1)} = \sum_{i,j_1, \ldots, j_{d_y}=1}^{N} c_{0,0,\beta} M_{\mu i j_1} W_{i j_1 j_2} M_{j_2 j_3}^+ \cdots M_{j_{d_y} j_{d_y+1}}^+,$$  \hspace{1cm} (S142)

$$T_{\mu\alpha\beta}^{(d_x+d_y+1)} = \sum_{i,j_1, \ldots, j_{d_y}=1}^{N} c_{i d_x d_y} M_{\mu i_0} M_{i_0 a_1}^+ \cdots M_{i_{d_x} a_{d_x}}^+ W_{i j_1} \cdots W_{i j_{d_y}} M_{j_{d_y} j_{d_y+1}}^+ \cdots M_{j_{d_y} j_{d_y+1}}^+,$$  \hspace{1cm} (S143)

for some real coefficients $c_{i d_x d_y}$ with $i \in \{1, \ldots, N\}$ 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.,

$$h_i(x_i, y_i) = \sum_{d_x = 0}^{\infty} \sum_{d_y = 0}^{\infty} c_{i, d_x d_y} x_i^{d_x} y_i^{d_y}, \quad i \in \{1, \ldots, N\},$$  \hspace{1cm} (S144)

where we have chosen to express the power series around $x_i = y_i = 0$ without loss of generality. The reduced dynamics is therefore

$$\dot{X}_\mu = \sum_{d_x, d_y = 0}^{\infty} \sum_{i=1}^{N} c_{i d_x d_y} M_{\mu i} \left( \sum_{\alpha=1}^{n} M_{i \alpha}^+ X_\alpha \right)^{d_x} \left( \sum_{\beta=1}^{n} \sum_{j=1}^{N} W_{ij} M_{j \beta}^+ X_\beta \right)^{d_y}.$$  \hspace{1cm} (S145)
The sum can be separated as
\[
\dot{X}_\mu = C_\mu + \sum_{d_x=1}^{\infty} \sum_{i=1}^{N} c_{id_x,0} M_{\mu i} \left( \sum_{\alpha=1}^{n} M^{+}_{\alpha\alpha} X_\alpha \right)^{d_x} + \sum_{d_y=1}^{\infty} \sum_{i=1}^{N} c_{id_y,0} M_{\mu i} \left( \sum_{j=1}^{n} W_{ij} M^{+}_{j\beta} X_\beta \right)^{d_y}
\]
\[
+ \sum_{d_x,d_y=1}^{\infty} \sum_{i=1}^{N} c_{id_x,d_y} M_{\mu i} \left( \sum_{\alpha=1}^{n} M^{+}_{\alpha\alpha} X_\alpha \right)^{d_x} \left( \sum_{j=1}^{N} W_{ij} M^{+}_{j\beta} X_\beta \right)^{d_y},
\]
where we have defined \( C_\mu = \sum_{i=1}^{N} M_{\mu i} c_{i00} \). Expanding the exponents and introducing the multi-indices directly provide the desired result.

**Remark S49.**

1. As explained in Section 1.1 (p.3) of Ref. [193], 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 S55, 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., \( T_1(2,3)(4) \) (\( d_x = 2, d_y = 1 \)) and \( T_2(2,3,4) \) (\( d_x = 1, d_y = 2 \)) are elements of two different tensors. Finally, when it’s clear in the context, if a multi-index is a singleton, then we remove the parentheses, e.g., \( T_{\mu i}(\epsilon) \) becomes \( T_{\mu i} \).
3. The coefficients \( c_{id_x,d_y} \) can be chosen as the ones of the Taylor series of \( h_t \) for all \( i \).
4. For the sake of simplicity, let us consider the case where \( c_{i01} = 1 \) for all \( i \). We observe that \( W^{(2)} = MWM^+ \) appears in the reduced dynamics, which can be viewed as the reduced weight matrix. From Theorem S37, it is also the unique solution to the compatibility equation [122] \( WM = MW \) when rank \( M = n \) and it is the least-square optimal solution to the problem \( \|WM - MW\|_F^2 \) with \( W \) as the optimization variable. Remember from Ref. [122] 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 \( X = MX \), \( \dot{X} = MWMX = WMX = WX \) where one can reasonably choose \( W = W^{(2)} \) 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 \( \bar{\cdot} \) is complex conjugation and \( r_i : \mathbb{C}^4 \rightarrow \mathbb{C} \) is a holomorphic function (and thus analytic):
\[
r_i(x_i, y_i, \bar{x}_i, \bar{y}_i) = \sum_{d_x=0}^{\infty} \sum_{d_y=0}^{\infty} \sum_{d_x=0}^{\infty} \sum_{d_y=0}^{\infty} c_{id_x,d_y} x_i^{d_x} y_i^{d_y} \bar{x}_i^{d_x} \bar{y}_i^{d_y}, \quad i \in \{1, \ldots, N\}.
\]
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 S55).
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. [122], 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 \( h_t(x_i, y_i) \). 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 \( K \) 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 \( c_{id_x,d_y} \)) is thus replaced by a hypergraph \( \mathcal{H} \) [193–195] with \( n \) vertices [see Fig. 2d of the paper], defined from the tensors \( D^{(d_x + 1)}, \mathcal{V}^{(d_y + 1)}, \) 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 S50.** A hypergraph is a triple \( \mathcal{H} = (\Upsilon, \Xi, \Omega) \), where
- \( \Upsilon = \{1, \ldots, n\} \) is the set of vertices;
- \( \Xi \) 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 \( \Upsilon \)), \( T \) is the tail of the hyperarc (a \( n_T \)-tuple with elements in \( \Upsilon \)), 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;
- \( \Omega \) is a function that assigns a real value to the hyperarcs.
Remark S51.

- The latter definition is a generalization of hypergraphs [194] and of directed hypergraphs as defined in Ref. [195], 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. [195]) is directed from the tail $T$ to the head $H$. As a consequence, in the tensor notation $T_{i\alpha\beta}^{(d_i+d_j+1)}$, the index $\mu$ and the multi-index $\alpha$ are part of the head while $\beta$ is part of the tail of the hyperarc. Thus, $T_{(i,2)}^{(1,3)}$ ($n_T = 1$) is a forward hyperarc $(1,2)$, $(3)$ while $T_{(1,2)}^{(2,3)}$ ($n_H = 1$) is a backward hyperarc $(1,2,3)$. Note that the tensor $\mathcal{Y}_{i\alpha\beta}^{(d_i+1)}$ with elements in Eq. (S142) always form backward hyperarcs (from $\beta$ to $\mu$) since $d_\mu = 0$, while the tensor with elements in Eq. (S143) can be any type of hyperarc (with $\mu$ always belonging to the head).

In the example of the paper for the epidemiological dynamics, Eq. (6) is a forward hyperarc (from $\kappa$ to $\mu$).

We now derive two key consequences of Proposition S48. First, Proposition S48 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 S52.** If $h_i(x_i, y_i)$ is a polynomial of total degree $\delta$ in $x_i$ and $y_i$ for all $i \in \{1, \ldots, N\}$ and condition (2) of Assumptions S42 holds, then the reduced dynamics has a polynomial vector field of total degree $\delta$ with interactions of maximal order $\delta + 1$.

**Proof.** Since any polynomial is analytic, condition (1) of Assumptions S42 is satisfied. Then, by Proposition S48, the reduced dynamics is given by Eqs. (S140-S143). In the following, the conclusions are valid for all $i \in \{1, \ldots, N\}$. Let $\mathcal{S}_i = \{c_{\mu_1 \cdots \mu_d}^{d_i, d_y} = 0\}$ 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 set of nonzero coefficients $\mathcal{F}_i \subset \mathcal{S}_i$ describing a polynomial vector field for the reduced dynamics. Consider any coefficient $c_{\mu_1 \cdots \mu_d}^{d_i, d_y} \in \mathcal{F}_i$ such that $d_\mu + d_y = \delta$, the total degree of the polynomial $h_i$. Then, at least one of the tensors $\mathcal{D}^{(d_i+1)}$, $\mathcal{W}^{(d_i+1)}$, $\mathcal{T}^{(d_i+d_j+1)}$, with elements in Eqs. (S141-S143), have the highest possible order $\delta + 1$. Moreover, there will be at least one monomial term $X_{\alpha_1} \ldots X_{\alpha_{d_i}} X_{\beta_1} \ldots X_{\beta_{d_y}}$, or $X_{\alpha_1} \ldots X_{\alpha_{d_i}} X_{\beta_1} \ldots X_{\beta_{d_y}}$ in Eq. (S140) that is of maximal degree $\delta$, which means that reduced dynamics has a polynomial vector field of total degree $\delta$.

Second, the tensors describe Proposition S48 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 S53.** Let $s : \mathcal{V} \to \mathcal{Y}$ be a surjection where $\mathcal{V} = \{1, \ldots, N\}$ and $\mathcal{Y} = \{1, \ldots, n\}$ are the vertex sets of the complete and reduced system respectively. If Assumptions S42 hold, the reduction matrix $M$ has elements $M_{\mu i} = m_{\mu i} \delta_{s(i)}$ with $m_{\mu i} \in \mathbb{R}$ for all $\mu, i$, 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 $\mu \in \mathcal{Y}$ and $i \in \mathcal{V}$, $M_{\mu i}^+ = m_{\mu i} \delta_{s(i)} / q_\mu$, where $q_\mu = \sum_{i=1}^N m_{\mu i}^2 \delta_{s(i)}$. Substituting $M$ and $M^+$ in Eqs. (S141-S143) yields

\[
\mathcal{D}^{(d_i+1)} = \frac{1}{q_\alpha} \sum_{i=1}^N c_{i \alpha d_\mu} m_{\mu i} \delta_{s(i)} \delta_s \delta_{s(i)} \alpha_1 \ldots \delta_{s(i)} \alpha_d, \\
\mathcal{W}^{(d_i+1)} = \frac{1}{q_\beta} \sum_{i,j_1, \ldots, j_{d_y} = 1}^N c_{i \beta d_y} m_{\mu i} \delta_{s(i)} W_{ij_1} \ldots W_{ij_{d_y}} \delta_{s(j_1)} \beta_1 \ldots \delta_{s(j_{d_y})} \beta_{d_y}, \\
\mathcal{T}^{(d_i+d_j+1)} = \frac{1}{q_\alpha q_\beta} \sum_{i,j_1, \ldots, j_{d_y} = 1}^N c_{i \alpha \beta d_y} m_{\mu i} \delta_{s(i)} \delta_{s(i)} \alpha_1 \ldots \delta_{s(i)} \alpha_{d_i} W_{ij_1} \ldots W_{ij_{d_y}} m_{\beta j_1} \beta_1 \ldots m_{\beta j_{d_y}} \beta_{d_y} \delta_{s(j_1)} \beta_1 \ldots \delta_{s(j_{d_y})} \beta_{d_y},
\]

where $q_\gamma = q_{\gamma_1} \cdots q_{\gamma_d}$ and $m_{\mu \alpha \beta i \cdots \alpha_d i} = m_{\mu i} m_{\alpha_1 i} \cdots m_{\alpha_d i}$. For $\mathcal{D}^{(d_i+1)}$ and any dependence of $h_i$ over $y_i$, it is straightforward to observe that the only non-zero elements are such that $\mu = \alpha_1 = \cdots = \alpha_{d_i}$. The tensor can therefore be mapped to a $n \times n$ diagonal matrix. Henceforth, we only consider $\mathcal{W}^{(d_i+1)}$ and $\mathcal{T}^{(d_i+d_j+1)}$.

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. (S144), i.e.,

\[
h_i(x_i, y_i) = \sum_{d_y = 0}^\infty c_{i d_y} x_i^d y_i, \quad i \in \{1, \ldots, N\}.
\]
Proposition S48 thus implies that

\[ W_{\mu \beta}^{(2)} = \frac{1}{q_{\beta}} \sum_{i,j=1}^{N} c_{i01} m_{\mu i} m_{\beta j} \delta_{\mu s(i)} W_{ij} \delta_{s(j) \beta} \] and \( T_{\mu \alpha \beta}^{(d+2)} = \frac{1}{q_{\alpha \beta}} \sum_{i,j=1}^{N} c_{i d+1} m_{\mu \alpha} m_{\beta j} \delta_{\mu s(i)} \delta_{s(i) \alpha_1} \ldots \delta_{s(i) \alpha_{d+2}} W_{ij} m_{\beta j} \delta_{s(j) \beta} . \]

Clearly, \( W_{\mu \beta}^{(2)} \) is a matrix and the nonzero elements of \( T_{\mu \alpha \beta}^{(d+2)} \) are for \( \mu = \alpha_1 = \ldots = \alpha_{d+2} \) (there are at most \( n^2 \) of them), which means that it can be mapped to the set \( x \times \n \) of vertices. Hence, there are solely pairwise interactions in the least-square reduced dynamics.

If \( d_q > 1 \) (i.e., for a nonlinear dependency of \( h_i \) over \( y_i \)), a simple example suffices to prove the last statement. Let \( d_q = 2 \), \( \mathcal{V} = \{1, 2, 3, 4, 5\} \), \( \mathcal{Y} = \{1, 2, 3\} \), and \( s(1) = 1 \), \( s(2) = 2 \), \( s(3) = 3 \), \( s(4) = 4 \), \( s(5) = 5 \). Moreover, consider that \( W_{12}, W_{13}, m_{12}, m_{23}, m_{35} \) are not equal to zero. Then, Proposition S48 gives

\[ W_{\mu \beta}^{(3)} = \frac{1}{q_{\beta_1 \beta_2}} \sum_{i=1}^{N} c_{i02} m_{\mu i} \delta_{\mu s(i)} \left( \sum_{j,k=1}^{N} m_{\beta_1 j} m_{\beta_2 k} W_{ij} W_{ik} \delta_{s(j) \beta_1} \delta_{s(k) \beta_2} \right) . \]

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 parenthesis of the last equation, there is a nonzero term \( m_{12} m_{23} W_{12} W_{13} \delta_{s(3) \beta_2} = m_{12} m_{23} W_{12} W_{13} \). Considering the whole equation for \( i = 5 \) and \( \mu = 3 \), there is a nonzero term \( c_{i02} m_{35} m_{12} m_{23} W_{12} W_{13} / q_{12} \). Hence, in this example, \( W_{3(1,2)} \neq 0 \) in general despite the fact that the observables are defined on disjoint sets of vertices.

\( \square \)

Remark S54. 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 \( X_i = h_i(x_i, y_i) \) 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 of \( x \) (without higher-order interactions), since \( x = M^{-1} X \).

Let’s now provide the details about the examples presented in Table III of the methods.

Example S55 (Emergence of higher-order interactions in typical models). Proposition S48 and Corollary S52 imply the following results in different fields of application.

1. QMF SIS dynamics [Eq. (S130) in Example S43]:

\[ \dot{X}_\mu = \sum_{\nu=1}^{n} (D_{\mu \nu} + W_{\mu \nu}) X_\nu + \sum_{\nu,\kappa=1}^{n} T_{\mu \nu \kappa} X_\nu X_\kappa, \] (S148)

where \( D = -MDM^+ \) with \( D = \text{diag}(d_1, \ldots, d_N) \), \( W = \gamma MWM^+ \), and

\[ T_{\mu \nu \kappa} = -\gamma \sum_{i,j=1}^{N} M_{\mu i} M_{+i \nu} W_{ij} M_{+j \kappa}, \]

with \( \alpha = (\nu) = \nu \) and \( \beta = (\kappa) = \kappa \). Interestingly, for \( n = 1 \), one can find the exact solution to Eq. (S130) since it is a Bernouilli differential equation.

2. Microbial population dynamics [Eq. (S135) in Example S45]:

\[ \dot{X}_\mu = C_\mu + \sum_{\nu=1}^{n} D_{\mu \nu} X_\nu + \sum_{\nu,\kappa=1}^{n} (D_{\mu (\nu, \kappa)} + T_{\mu \nu \kappa}) X_\nu X_\kappa + \sum_{\nu,\kappa,\tau=1}^{n} D_{\mu (\nu, \kappa, \tau)} X_\nu X_\kappa X_\tau \] (S149)

where \( D = -dMM^+ \) and

\[ D_{\mu (\nu, \kappa)} = b \sum_{i=1}^{N} M_{\mu i} M_{+i \nu} M_{+i \kappa}^+, \quad T_{\mu \nu \kappa} = \gamma \sum_{i,j=1}^{N} M_{\mu i} M_{+i \nu} W_{ij} M_{+j \kappa}^+, \quad D_{\mu (\nu, \kappa, \tau)} = -c \sum_{i=1}^{N} M_{\mu i} M_{+i \nu} M_{+i \kappa}^+ M_{+i \tau}^+. \]

3. Kuramoto-Sakaguchi dynamics [Eq. (S137) in Example S46]:

\[ \dot{X}_\mu = \sum_{\nu=1}^{n} (D_{\mu \nu} + W_{\mu \nu}) X_\nu + \sum_{\nu,\kappa,\tau=1}^{n} T_{\mu (\nu, \kappa, \tau)} X_\nu X_\kappa X_\tau \] (S150)
where $\mathcal{D} = iMDM^+$ with $D = \text{diag}(\omega_1, ..., \omega_N)$, $W = \gamma e^{-i\alpha}MM^+$, and

$$T_{\mu(\nu,\kappa)} = -\gamma e^{i\alpha} \sum_{j,k=1}^{N} M_{j\mu}M_{j\mu}^+M_{k\nu}^+ W_{jk}M_{k\nu}^+,$$

with $\alpha = (\nu, \kappa), \beta = (\tau) = \tau$. In this case, the reduced variables $X_1, ..., X_n$ and the involved tensors are complex.

Remark S56.

- We found that there can be computational benefits to write the vector fields in terms of tensors (SI 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. [196] or Ref. [197]. See Ref. [158] for more references.

In phase reduction techniques [178], $dN$-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 [180, 198, 199] or, in other words, between microscopic observables (i.e., there is a phase for each oscillator, considered as the microscopic level, except in Ref. [200, 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 S39 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 S57.

1. The $N$-dimensional complete dynamics on a network defined by the real $N \times N$ weight matrix $W$ is

$$\dot{x} = g(x, y) = \begin{pmatrix} g_1(x_1, ..., x_N, y_1, ..., y_N) \\ \vdots \\ g_N(x_1, ..., x_N, y_1, ..., y_N) \end{pmatrix}, \quad (S151)$$

where $x : t \rightarrow \mathbb{R}^N$, $y = Wx$, and $g : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a continuously differentiable function.

2. The $n$-dimensional reduced dynamics ($n < N$) is the least-square optimal dynamics of Theorem S39, i.e.,

$$\dot{X} = Mg(M^+X, WM^+X). \quad (S152)$$

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 S42. 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 $n = N = \text{rank} M$. In this case, the “reduced” dynamics is not reduced anymore, but it is still a linear transformation of the complete dynamics.

Lemma S58. If conditions (1) and (2) in Assumptions S57 hold with $n = N = \text{rank} M$, 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 $\sigma_n$ decrease rapidly with $n$, so does the alignment error. In other words, the low-rank hypothesis of networks induces a low-dimension hypothesis of dynamical systems.
Theorem S59. If all conditions of Assumptions S57 hold, the alignment error in \( \mathbb{R}^n \) at \( x \in \mathbb{R}^N \) is upper-bounded as
\[
\mathcal{E}(x) \leq \frac{1}{\sqrt{n}} \left[ \| V_n^\top J_x(x', y')(I - V_n V_n^\top) x \| + \sigma_{n+1} \| V_n^\top J_y(x', y') \|_2 \| x \| \right],
\] (S153)
where \( y' = Wx' \) with \( x' \) being some point between \( x \) and \( V_n V_n^\top x \), \( \sigma_i \) 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:
\[
\frac{\mathcal{E}(x)}{\| x \|} \leq \frac{1}{\sqrt{n}} \left[ \alpha(x', y') + \sigma_{n+1} \beta(x', y') \right],
\] (S154)
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 S57, we have
\[
\mathcal{E}(x) = \frac{1}{\sqrt{n}} \| M [g(x, y) - g(\tilde{x}, \tilde{y})] \|,
\] (S155)
where \( y = Wx, \tilde{x} = Px, \) and \( \tilde{y} = WPx \) with \( P = M^+ M \). Let’s define the function
\[
u(x) = g(x, \ell(x)),
\] (S156)
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
\[
u(x) = u(\tilde{x}) + Du(x')(x - \tilde{x})
\] (S157)
for some \( x' \) between \( x \) and \( Px \) and where \( Du(x') \) is the total derivative of \( u \). From Eq. (S156) and the chain rule for the total derivative, we have (abusing the matrix notation)
\[
Du(x') = Dg(x', y') \frac{\partial g}{\partial x}(x', y') \frac{\partial \ell}{\partial x}(x') = J_x(x', y') + J_y(x', y')W,
\] (S158)
where \( y' = \ell(x') = Wx' \), the elements of the Jacobian matrices \( J_x(x', y') \), \( J_y(x', y') \) are respectively
\[
[J_x(x', y')]_{ij} = \frac{\partial g_i(x, y)}{\partial x_j}(x, y), \quad [J_y(x', y')]_{ij} = \frac{\partial g_i(x, y)}{\partial y_j}(x', y'),
\] (S159)
and we have used the fact that \( W \) is the Jacobian matrix of \( \ell \). The Taylor expansion (S157) 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 \( y' = Wx' \):
\[
g(x, y) = g(\tilde{x}, \tilde{y}) + J_x(x', y')(x - \tilde{x}) + J_y(x', y')W(x - \tilde{x}).
\] (S160)
The alignment error becomes
\[
\mathcal{E}(x) = \frac{1}{\sqrt{n}} \| M [J_x(x', y')(I - P)x + J_y(x', y')W(I - P)x] \|
\] (S161)
and the triangle inequality gives
\[
\mathcal{E}(x) \leq \frac{1}{\sqrt{n}} \left[ \| MJ_x(x', y')(I - P)x \| + \| MJ_y(x', y')W(I - P)x \| \right].
\] (S162)
Moreover, the induced spectral norm for the second term yields
\[
\mathcal{E}(x) \leq \frac{1}{\sqrt{n}} \left[ \| MJ_x(x', y')(I - P)x \| + \| MJ_y(x', y')W(I - P) \|_2 \| x \| \right]
\] (S163)
and the submultiplicativity of the spectral norm implies
\[
\mathcal{E}(x) \leq \frac{1}{\sqrt{n}} \left[ \| MJ_x(x', y')(I - P)x \| + \| W(I - P) \|_2 \| MJ_y(x', y') \|_2 \| x \| \right].
\] (S164)
From condition (3) of Assumption S57, we have \( M = V_n^\top \) which is, by Theorem S11, the optimal solution to the minimization of \( \| W(I - P) \|_2 \) with error \( \sigma_{n+1} \) (square root of the problem (P2)) and the error for the spectral norm
The second inequality is deduced as follows:

\[
\frac{\mathcal{E}(x)}{\|x\|} \leq \frac{1}{\sqrt{n}} \left[ \|V_n^T J_x(x', y')(I - V_n V_n^T)\|_2 + \sigma_{n+1} \|V_n^T J_y(x', y')\|_2 \right] \\
\leq \frac{1}{\sqrt{n}} \left[ \|V_n^T J_x(x', y')\|_2 + \sigma_{n+1} \|V_n^T J_y(x', y')\|_2 \right] \\
\leq \frac{1}{\sqrt{n}} \left[ \|J_x(x', y')\|_2 + \sigma_{n+1} \|J_y(x', y')\|_2 \right],
\]

where we have used successively the submultiplicativity of the spectral norm and identities \(\|I - V_n V_n^T\|_2 = 1\), \(\|V_n^T J_x(x', y')\|_2 = \|J_x(x', y')\|_2\) and \(\|V_n^T J_y(x', y')\|_2 = \|J_y(x', y')\|_2\). 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 S60.**

- The dynamics used in the paper have the less general form (compared to condition (1) in Assumptions S57)
  \[
  \dot{x}_i = h_i(x_i, y_i), \quad i \in \{1, ..., N\},
  \]
  where, for all \(i, x_i : t \rightarrow \mathbb{R}^N\), \(y_i = \sum_{j=1}^{N} W_{ij} x_j\) and \(h_i : \mathbb{R}^2 \rightarrow \mathbb{R}\). 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 \(\sigma_{n+1}\) is not negligible when \(n = 1\). According to our analysis, we therefore do not expect one-dimensional reductions \([175, 201, 202]\) to yield accurate results in general, which is consistent with numerical observations made in previous studies \([122, 201, 203–205]\). Some very simple synthetic networks, however, such as those generated by the Erdős-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) \(\sigma_{n+1}\) 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 \(\sigma_{n+1}\) in real networks, the factors \(\alpha(x', y')\) and \(\beta(x', y')\) do not necessarily decrease as \(n\) increases.

As a byproduct of the last theorem, the fact that the term \(\|W(I - M^+ M)\|_2\) appears in the upper bound in Eq. (S164) of the alignment error suggests a reasonable choice of reduction matrix, \(M = V_n^T\), which minimizes \(\|W(I - M^+ M)\|_2\) from Theorem S11. Of course, this doesn’t mean that it is the reduction matrix that minimizes the alignment error in \(\mathbb{R}^n\) (which is a hard 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 dimension or in image, perfect alignment of the complete and reduced vector fields, as shown in the following corollary.

**Corollary S61.** If all conditions of Assumptions S57 hold, \(J_x(x', y') = aI\) for some real constant \(a\), and \(n = \text{rank} W\), 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 S59 for any \(M\):

\[
\|MJ_x(x', y')(I - P)x\| = a \|(M - MM^+ M)x\| = 0,
\]

since \(MM^+ M = M\) according to the defining properties of the Moore-Penrose pseudoinverse. Finally, if \(n = \text{rank} W\), then \(\sigma_{n+1} = 0\), which cancels out the second term of the bound.

Let \(s\) be a vector of \(N\) functions \(s_i : \mathbb{R} \rightarrow \mathbb{R}\). Let \(W\) be a \(N \times N\) matrix of rank \(r < N\) with compact SVD \(U_r \Sigma_r V_r^T\). If \(M = V_r^T\), then by Corollary S61, the dynamics

\[
\dot{x} = -d x + s(Wx)
\]

can be exactly reduced to the \(r\)-dimensional reduced dynamics

\[
\dot{X} = -dX + V_r^T s(U_r \Sigma_r X), \quad X = V_r^T x.
\]
Example S62. A noteworthy example of dynamics of the form (S170) is the RNN defined by Eqs. (S139). Therefore, when \( \text{rank } W = r < N \) and \( d_i = d \) for all \( i \in \{1, \ldots, N\} \), the RNN exactly reduces to the \( r \)-dimensional dynamics

\[
\dot{X} = -dX + V_r^\top \tanh(U_r \Sigma_r X + c), \quad X = V_r^\top x, \tag{S172}
\]

where \( U_r, \Sigma_r, V_r^\top \) form the compact SVD of the neural network \( W \).

The RNN used in reservoir computing \([97]\) also involves a dynamics of the form (S139) (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. Yet, we can expect it to have a low effective rank, given the result in the paper. By shrinking the singular values (with optimal shrinkage \([27]\) 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 solely 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 \( W_k \).

Example S63. The Wilson-Cowan dynamics in Eq. (S126) with \( a = 0 \) and \( d_j = d \) for all \( j \in \{1, \ldots, N\} \) is essentially equivalent, from a mathematical perspective, to the RNN of the last example. It can thus be exactly reduced to the dynamics

\[
\dot{X} = -dX + V_n^\top [b(\gamma U_n \Sigma_n X - c)], \quad X = V_n^\top x. \tag{S173}
\]

In Fig. S8, we illustrate this result for a real connectome by comparing the global observable at equilibrium (see Section III F) of the complete and reduced dynamics \( \dot{X} = -dX + V_n^\top [b(\gamma U_n \Sigma_n X - c)] \) with \( X = V_n^\top x \) and different values of \( n \).

Example S64. The threshold-linear model in Eq. (S132) with \( \text{rank } W = r < N \) can also be exactly reduced (despite the discontinuity in the vector field) to the \( r \)-dimensional reduced dynamics

\[
\dot{X} = -X + V_r^\top [U_r \Sigma_r X + b]_+^+, \quad X = V_r^\top x, \tag{S174}
\]

where \( X = V_r^\top x \) and \( U_r, \Sigma_r, V_r^\top \) form the compact SVD of the neural network \( W \). To apply Corollary S61, one can simply replace \([ ]^+\) by the softplus function to satisfy the condition 1 of Assumptions S57.

### E. Computation of the upper bound on the alignment error

The bound in Theorem S59 depends on some real point \( x' \) which is unknown \textit{a priori}. Yet, according to Eqs. (S157-S160), it is possible to find \( x' \) analytically (sometimes exactly) or numerically from

\[
[G_x(x') + G_y(x') W] (I - P)x = u(x) - u(Px), \tag{S175}
\]

where \( G_x(x') = J_x(x', y') \), and \( G_y(x', y') = J_y(x', y') \). Below, we give four examples, one for each dynamics used in the paper to produce Fig. 3, from the simplest to the more complex case.
**Example S65** (Epidemiological). For the QMF SIS dynamics in Eq. (S130), we can exactly find \( x' \). We have
\[
\begin{align*}
    u(x) &= -Dx + \gamma (1 - x) \circ Wx \\
    G_x(x') &= -D - \gamma \text{diag}(Wx') \\
    G_y(x') &= \gamma [I - \text{diag}(x')],
\end{align*}
\]
where \( D = \text{diag}(d_1, ..., d_N) \). By substituting the expressions above in Eq. (S175) and by canceling some terms, we have
\[
Wx' \circ \chi + x' \circ W\chi = x \circ Wx - Px \circ WPx.
\]
where \( \chi = (I - P)x \). The commutativity of the Hadamard product implies
\[
\chi \circ Wx' + W\chi \circ x' = x \circ Wx - Px \circ WPx,
\]
which can be written as a linear equation in \( x' \), i.e.,
\[
[\text{diag}(\chi)W + \text{diag}(W\chi)]x' = x \circ Wx - Px \circ WPx,
\]
If the matrix \( \text{diag}(\chi)W + \text{diag}(W\chi) \) is invertible (which is true in general), then the unique solution to the linear system is
\[
x' = [\text{diag}(\chi)W + \text{diag}(W\chi)]^{-1}(x \circ Wx - Px \circ WPx).
\]
In rare cases, if the matrix \( \text{diag}(\chi)W + \text{diag}(W\chi) \) is singular, then one can use the least-square optimal solution by using the pseudo-inverse. Therefore, using Eq. (S182), one can compute exactly the upper bound on the alignment error for the QMF SIS. In Fig. 3a, 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 \( \gamma \) is sampled from a uniform probability density function between 0.01 and 4. In this parameter region, there is a bifurcation for the global observable defined in Sec. III F (see Fig. S11a).

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 S66** (RNN). Another way to write the RNN (with no current) is
\[
\dot{x}_i = -d_i x_i + \tanh(\gamma \sum_{j=1}^{N} W_{ij} x_j) = -d_i x_i + 2\gamma \sum_{j=1}^{N} W_{ij} x_j - 1.
\]
For the RNN dynamics, we have
\[
\begin{align*}
    u(x) &= -Dx + \tanh(\gamma Wx) = -Dx + 2\gamma (2\gamma Wx) - 1 \\
    G_x(x') &= -D \\
    G_y(x') &= 4\gamma \text{diag} \left[ S(2\gamma Wx')[1 - S(2\gamma Wx')] \right]
\end{align*}
\]
where \( D = \text{diag}(d_1, ..., d_N) \), \( S \) 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 \( S(2\gamma Wx') \) so we won’t have to look for \( x' \). By substituting the expressions above in Eq. (S175) and by canceling some terms, we get
\[
S(2\gamma Wx')[1 - S(2\gamma Wx')] = \frac{1}{2\gamma} \text{diag}[W(I - P)x]^{-1}[S(2\gamma Wx) - S(2\gamma WPx)]
\]
which can be directly substituted into \( G_y(x') \) to compute the upper bound. In Fig. 3c, we compute the bound for the learned network mouse-control-1-model.npz from Ref. [206]. 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 1/0.625 and standard deviation 3/0.625 and the coupling constant \( \gamma \) is sampled from a uniform probability density function between 0.1/0.625 and 0.001. In this parameter region, inspired by Ref. [206], the parameter region includes the chaotic regime of the RNN. See the script simulations/bifurcations_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 S67 (Neuronal). For the Wilson-Cowan dynamics, we have
\begin{align}
    u(x) &= -Dx + (1 - ax) \circ S[b(\gamma Wx - c)] \\
    G_x(x') &= -D - a \text{diag}(S[b(\gamma Wx' - c)]) \\
    G_y(x') &= \gamma [I - \text{diag}(x')] \text{diag}(S[b(\gamma Wx' - c)](1 - S[b(\gamma Wx' - c)]))
\end{align}
\hspace{1cm} (S188) \hspace{1cm} (S189) \hspace{1cm} (S190)

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. (S175) 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 \( \gamma \) is sampled from a uniform probability density function between 0.01 and 1. In Fig S3b, 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 S67 that helps solve Eq. (S175) in \( x' \).

Example S68 (Microbial). For the microbial population dynamics defined in Eq. (S135), we have
\begin{align}
    u(x) &= a - dx + b x \circ x - c x \circ x + \gamma x \circ Wx \\
    G_x(x') &= -dI + \text{diag}(2bx' - 3cx' \circ x' + \gamma Wx') \\
    G_y(x') &= \gamma \text{diag}(x').
\end{align}
\hspace{1cm} (S191) \hspace{1cm} (S192) \hspace{1cm} (S193)

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 \( x \approx 0 \) (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 \( n \in \{1,50,100,150,200,250,296\} \)).
In matrix form, it is easily shown that the system of equations to solve is

$$A(x' \circ x') + Bx' - C = 0,$$

where $\chi = (I - P)x$, $D_v = \text{diag}(v_1, ..., v_N)$, and

$$A = -3cD_\chi,$$

$$B = 2bD_\chi + \gamma D_W \chi + \gamma D_\chi W,$$

$$C = b[x \circ x - Px \circ Px] - c[x \circ x - Px \circ Px] + \gamma [x \circ (Wx) - (Px) \circ (WPx)].$$

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 $\gamma D_\chi Wx'$ nor the quadratic term can be neglected. Moreover, for the parameters $a = 5$, $b = 13$, $c = 10/3$, $d = 30$ (with corresponding values $F = 5$, $B = 3$, $C = 3$, $K = 10$ in Ref. [173]), $\gamma \in [0.5, 3]$ and the human gut microbiome network [173, 207] (parameters that yields a hysteresis for the global observables equilibrium points as in Fig. S11c), 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, and so we set $x_i \mapsto x_i/30$. Thus, with $t \mapsto 20\sigma_1 t$, we get the differential equations

$$\frac{dx_i}{dt} = a - dx_i + bx_i^2 - cx_i^3 + \gamma x_i \sum_{j=1}^{N} W_{ij} x_j,$$

where the parameters are redefined such that $a \mapsto a/T \approx 5 \times 10^{-5}$, $d \mapsto d/T \approx 0.01$, $b \mapsto bd/T \approx 0.1$, $c \mapsto cd^2/T \approx 0.9$, $\gamma \mapsto \gamma d/20 \in [0.5, 4.5]$, and $W = W/\sigma_1$. In Fig. 3d, we use the second method in Example S67 ($x'$ 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. 3d. 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 $\gamma$ is sampled from a uniform probability density function between 0.1 and 5.
F. Global bifurcation diagram

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 \( V = (v_1 \ldots v_N) \) 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 \( X_\mu \) (even the leading one, i.e., \( X_1 \) related to \( \sigma_1 \)) will have equilibrium points below 0, which are often harder to interpret. One way to get more positive values without using any approximation (e.g., nonnegative matrix factorization) 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

\[
D_\pm = \text{diag} \left[ s(\sum_i (v_1)_i), \ldots, s(\sum_i (v_N)_i) \right],
\]

where \( s \) is defined such that

\[
s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0. \end{cases}
\]

Since \( D_\pm \) is diagonal, it commutes with any diagonal matrices. Moreover, \( D_+ D_- = I \). Therefore,

\[
W = U \Sigma V^T = U \Sigma D_\pm D_\pm V^T = U D_\pm \Sigma D_\pm V^T := U \Sigma V^T.
\]

To get an approximate reduced dynamics of dimension \( n \), we use the truncated SVD \( U_n \Sigma_n V_n^T \), where \( \Sigma_n := \text{diag}(\sigma_1, \ldots, \sigma_n) \) and the \( N \times n \) truncated singular vector matrices are

\[
U_n := (u_1 \ldots u_n) \text{ and } V_n := (v_1 \ldots v_n^\prime).
\]

After integrating the reduced dynamics, we can compute the global observable

\[
\mathcal{X} = \sum_{\mu=1}^n \left( \sum_{i=1}^{n} \frac{\ell_\mu}{\sum_{i=1}^{n} \sum_{\mu=1}^{n} \ell_\mu M_{\mu i}} \right) X_\mu = \sum_{i=1}^{N} \left( \sum_{\mu=1}^{n} \frac{\ell_\mu M_{\mu i}}{\sum_{i=1}^{n} \sum_{\mu=1}^{n} \ell_\mu M_{\mu i}} \right) x_i,
\]

where \( M = V_n^T \). If \( \ell = (1 \; 0 \; 0) \) for instance, then \( \mathcal{X} \) is the leading right-singular-vector observable, which corresponds, when \((v_1)_i \geq 0\) for all \(i\), to the average activity weighted by the vertices’ hub centrality (cf. FIG. S5):

\[
\mathcal{X} = \sum_{i=1}^{N} \frac{(v_1)_i}{\sum_i (v_1)_i} x_i.
\]

In Fig. S11, we illustrate the bifurcation diagrams of the latter observable for three dynamics. For unsigned networks, such as Fig. S11 (a-b), the observable \( \mathcal{X} \) is between 0 and 1 which is guaranteed by the Perron-Frobenius theorem [68, Theorem 38]. Otherwise, the observable might \( \mathcal{X} \) can be negative, but it can still be insightful in terms of emergent phenomena as shown in the bifurcation diagrams in Fig. S11 (c) where there is an explosive transition.

Fig. S11: Stable bifurcation branches of the leading right-singular-vector observable of three different dynamics. The blue curves are the equilibrium points of the high-dimensional dynamics while the orange curves are the ones of the reduced dynamics. a, QMF SIS dynamics on a undirected network of high-school contacts with infection rate \( \gamma \). The complete dynamics has dimension \( N = 327 \) and the reduced dynamics has dimension \( n = 1 \) and matrix \( D = I \). b, Excitatory Wilson-Cowan dynamics on the weighted and directed connectome of \textit{Ciona intestinalis} with coupling \( \gamma \), \( N = 213 \), \( n = 20 \), \( a = 0.1 \), \( b = 1 \), \( c = 3 \), \( D = I \). c, Population dynamics on a signed, weighted, and directed gut microbiome network [173] with coupling \( \gamma \), \( N = 838 \), \( n = 76 \), \( a = 5 \), \( b = 13 \), \( c = 10/3 \), \( D = 30I \). We choose \( n = 76 \) (essentially, the elbow) by looking at the alignment error profile of the microbial population dynamics in Fig. 3.
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 SV: 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: $N = 500$, $x_i \sim U(0, 1)$, $\theta_i, \alpha \sim U(0, 2\pi)$, $D_{ij} \sim U(0, 1)$, $W_{ij} \sim U[-1, 1]$. 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.

| $n$ | Reduced vector field | Lotka-Volterra | QMF SIS | Kuramoto-Sakaguchi |
|-----|---------------------|---------------|---------|-------------------|
| 1   | Exact $M \circ h$   | $3.2 \times 10^{-4}$ | $3.3 \times 10^{-3}$ | $9.4 \times 10^{-3}$ |
|     | Unspecified $M \circ h \circ M^+$ | $3.3 \times 10^{-4}$ | $3.1 \times 10^{-3}$ | $9.0 \times 10^{-3}$ |
|     | Tensor form         | $1.0 \times 10^{-5}$ | $2.0 \times 10^{-5}$ | $5.6 \times 10^{-5}$ |
| 10  | Exact $M \circ h$   | $3.6 \times 10^{-4}$ | $3.4 \times 10^{-3}$ | $9.2 \times 10^{-3}$ |
|     | Unspecified $M \circ h \circ M^+$ | $3.0 \times 10^{-4}$ | $2.8 \times 10^{-3}$ | $8.9 \times 10^{-3}$ |
|     | Tensor form         | $2.0 \times 10^{-5}$ | $2.9 \times 10^{-5}$ | $9.2 \times 10^{-5}$ |
| 100 | Exact $M \circ h$   | $5.4 \times 10^{-4}$ | $3.4 \times 10^{-3}$ | $1.0 \times 10^{-2}$ * |
|     | Unspecified $M \circ h \circ M^+$ | $4.6 \times 10^{-4}$ | $3.3 \times 10^{-3}$ | $1.1 \times 10^{-2}$ * |
|     | Tensor form         | $8.6 \times 10^{-4}$ | $9.0 \times 10^{-4}$ | $1.2 \times 10^{0}$ * |

*Computed with 10 samples instead of 500.

As shown in Table SV, 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.

[1] E. Schmidt, Math. Ann. 63, 433 (1907).
[2] C. Eckart and G. Young, Psychometrika 1, 211 (1936).
[3] G. W. Stewart, 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, Pattern Recognit. 14, 375 (1981).
[6] H. Hotelling, J. Educ. Psych. 24, 417 (1933).
[7] H. Hotelling, J. Educ. Psych. 24, 498 (1933).
[176] Y. Kuramoto, in *International Symposium on Mathematical Problems in Theoretical Physics* (1975) p. 420.
[177] H. Sakaguchi and Y. Kuramoto, *Prog. Theor. Phys.* **76**, 576 (1986).
[178] B. Pietras and A. Daffertshofer, *Phys. Rep.* **819**, 1 (2019).
[179] K. Wiesenfeld, P. Colet, and S. H. Strogatz, *Phys. Rev. Lett.* **76**, 404 (1996).
[180] M. H. Matheny, J. Emenheiser, 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, *Science* **363**, 1057 (2019).
[181] E. M. Izhikevich, *Dynamical Systems in Neuroscience* (MIT Press, 2007).
[182] A. T. Winfree, *J. Theoret. Biol.* **16**, 15 (1967).
[183] G. B. Ermentrout and N. Kopell, *SIAM J. Appl. Math.* **46**, 233 (1986).
[184] K. I. Funahashi and Y. Nakamura, *Neural Netw.* **6**, 801 (1993).
[185] R. D. Beer, *Robot. Auton. Syst.* **20**, 257 (1997).
[186] M. A. Cohen and S. Grossberg, *IEEE Trans. Syst. Man Cybern. SMC-13*, 815 (1983).
[187] S. Grossberg, *Neural Netw.* **1**, 17 (1988).
[188] J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3088 (1984).
[189] M. W. Hirsch, *Neural Netw.* **2**, 331 (1989).
[190] R. Kim, Y. Li, and T. J. Sejnowski, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22811 (2019).
[191] R. Kim and T. J. Sejnowski, *Nat. Neurosci.* **24**, 129 (2021).
[192] J. Hanson and M. Raginsky, *Proc. Mach. Learn. Res.* **120**, 1 (2020).
[193] L. Qi and Z. Luo, *Tensor analysis* (SIAM, 2017).
[194] C. Berge, *Hypergraphs* (North-Holland, 1989).
[195] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen, *Discret. Appl. Math.* **42**, 177 (1993).
[196] G. Ferraz de Arruda, M. Tizzani, and Y. Moreno, *Commun. Phys.* **4**, 24 (2021).
[197] R. Mulas, C. Kuehn, and J. Jost, *Phys. Rev. E* **101**, 062313 (2020).
[198] P. Ashwin and A. Rodrigues, *Physica* **325**, 14 (2016).
[199] I. León and D. Pazó, *Phys. Rev. E* **100**, 012211 (2019).
[200] T. Pereira, Nat. Commun. **13**, 4849 (2022).
[201] E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, *Phys. Rev. X* **9**, 011042 (2019).
[202] P. Kundu, H. Kori, and N. Masuda, *Phys. Rev. E* **105**, 024305 (2022).
[203] C. Tu, J. Grilli, F. Schuessler, and S. Suweis, *Phys. Rev. E* **95**, 062307 (2017).
[204] J. Jiang, Z.-G. Huang, T. P. Seager, W. Lin, C. Grebogi, A. Hastings, and Y.-C. Lai, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E639 (2018).
[205] M. Véghé, V. Thibeault, P. Desrosiers, and A. Allard, arXiv:2206.11230 (2022).
[206] 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, *Neuron* **109**, 2556 (2021).
[207] R. Lim, J. J. T. Cabatbat, T. L. P. Martin, H. Kim, S. Kim, J. Sung, C. M. Ghim, and P. J. Kim, *Sci. Data* **7**, 1 (2020).