Article

Neuron

Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks

Highlights

- We study network models characterized by minimal connectivity structures
- For such models, low-dimensional dynamics can be directly inferred from connectivity
- Computations emerge from distributed and mixed representations
- Implementing specific tasks yields predictions linking connectivity and computations

Authors

Francesca Mastrogiuseppe, Srdjan Ostojic

Correspondence

srdjan.ostojic@ens.fr

In Brief

Neural recordings show that cortical computations rely on low-dimensional dynamics over distributed representations. How are these generated by the underlying connectivity? Mastrogiuseppe et al. use a theoretical approach to infer low-dimensional dynamics and computations from connectivity and produce predictions linking connectivity and functional properties of neurons.

Mastrogiuseppe & Ostojic, 2018, Neuron *99*, 609–623
August 8, 2018
https://doi.org/10.1016/j.neuron.2018.07.003

CellPress

Neuron
Article

Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks

Francesca Mastrogiuseppe[1,2] and Srdjan Ostojic[1,3,*]
[1]Laboratoire de Neurosciences Cognitives, INSERM U960, École Normale Supérieure - PSL Research University, 75005 Paris, France
[2]Laboratoire de Physique Statistique, CNRS UMR 8550, École Normale Supérieure - PSL Research University, 75005 Paris, France
[3]Lead Contact
*Correspondence: srdjan.ostojic@ens.fr
https://doi.org/10.1016/j.neuron.2018.07.003

SUMMARY

Large-scale neural recordings have established that the transformation of sensory stimuli into motor outputs relies on low-dimensional dynamics at the population level, while individual neurons exhibit complex selectivity. Understanding how low-dimensional computations on mixed, distributed representations emerge from the structure of the recurrent connectivity and inputs to cortical networks is a major challenge. Here, we study a class of recurrent network models in which the connectivity is a sum of a random part and a minimal, low-dimensional structure. We show that, in such networks, the dynamics are low dimensional and can be directly inferred from connectivity using a geometrical approach. We exploit this understanding to determine minimal connectivity required to implement specific computations and find that the dynamical range and computational capacity quickly increase with the dimensionality of the connectivity structure. This framework produces testable experimental predictions for the relationship between connectivity, low-dimensional dynamics, and computational features of recorded neurons.

INTRODUCTION

Understanding the relationship between synaptic connectivity, neural activity, and behavior is a central endeavor of neuroscience. Networks of neurons encode incoming stimuli in terms of electrical activity and transform this information into decisions and motor actions through synaptic interactions, thus implementing computations that underlie behavior. Reaching a simple, mechanistic grasp of the relation between connectivity, activity, and behavior is, however, highly challenging. Cortical networks, which are believed to constitute the fundamental computational units in the mammalian brain, consist of thousands of neurons that are highly inter-connected through recurrent synapses. Even if one were able to experimentally record the activity of every neuron and the strength of each synapse in a behaving animal, understanding the causal relationships between these quantities would remain a daunting challenge because an appropriate conceptual framework is currently lacking (Gao and Ganguli, 2015). Simplified, computational models of neural networks provide a test bed for developing such a framework. In computational models and trained artificial neural networks, the strengths of all synapses and the activity of all neurons are known, yet an understanding of the relation between connectivity, dynamics, and input-output computations has been achieved only in very specific cases (e.g., Hopfield (1982); Ben-Yishai et al. (1995); Wang (2002)).

One of the most popular and best-studied classes of network models is based on fully random recurrent connectivity (Sompolinsky et al., 1988; Brunel, 2000; van Vreeswijk and Sompolinsky, 1996). Such networks display internally generated irregular activity that closely resembles spontaneous cortical patterns recorded *in vivo* (Shadlen and Newsome, 1998). However, randomly connected recurrent networks display only very stereotyped responses to external inputs (Rajan et al., 2010), can implement only a limited range of input-output computations, and their spontaneous dynamics are typically high dimensional (Williamson et al., 2016). To implement more elaborate computations and low-dimensional dynamics, classical network models rely instead on highly structured connectivity, in which every neuron belongs to a distinct cluster and is selective to only one feature of the task (e.g., Wang (2002); Amit and Brunel (1997); Litwin-Kumar and Doiron (2012)). Actual cortical connectivity appears to be neither fully random nor fully structured (Harris and Mrsic-Flogel, 2013), and the activity of individual neurons displays a similar mixture of stereotypy and disorder (Rigotti et al., 2013; Mante et al., 2013; Churchland and Shenoy, 2007). To take these observations into account and implement general-purpose computations, a large variety of functional approaches have been developed for training recurrent networks and designing appropriate connectivity matrices (Hopfield, 1982; Jaeger and Haas, 2004; Maass et al., 2007; Sussillo and Abbott, 2009; Eliasmith and Anderson, 2004; Boerlin et al., 2013; Pascanu et al., 2013; Martens and Sutskever, 2011). A unified conceptual picture of how connectivity determines dynamics and

computations is, however, currently missing (Barak, 2017; Sussillo, 2014).

Remarkably, albeit developed independently and motivated by different goals, several of the functional approaches for designing connectivity appear to have reached similar solutions (Hopfield, 1982; Jaeger and Haas, 2004; Sussillo and Abbott, 2009; Eliasmith and Anderson, 2004; Boerlin et al., 2013), in which the implemented computations do not determine every single entry in the connectivity matrix but instead rely on a specific type of minimal, low-dimensional structure, so that in mathematical terms the obtained connectivity matrices are *low rank*. In classical Hopfield networks (Hopfield, 1982; Amit et al., 1985), a rank-one term is added to the connectivity matrix for every item to be memorized, and each of these terms fixes a single dimension, i.e., row/column combination, of the connectivity matrix. In echo state (Jaeger and Haas, 2004; Maass et al., 2007) and FORCE learning (Sussillo and Abbott, 2009), and similarly within the Neural Engineering Framework (Eliasmith and Anderson, 2004), computations are implemented through feedback loops from readout units to the bulk of the network. Each feedback loop is mathematically equivalent to adding a rank-one component and fixing a single row/column combination of the otherwise random connectivity matrix. In the predictive spiking theory (Boerlin et al., 2013), the requirement that information is represented efficiently leads again to a connectivity matrix with similar low-rank form. Taken together, the results of these studies suggest that a minimal, low-rank structure added on top of random recurrent connectivity may provide a general and unifying framework for implementing computations in recurrent networks.

Based on this observation, here we study a class of recurrent networks in which the connectivity is a sum of a structured, low-rank part and a random part. We show that in such networks, both spontaneous and stimulus-evoked activity are low-dimensional and can be predicted from the geometrical relationship between a small number of high-dimensional vectors that represent the connectivity structure and the feedforward inputs. This understanding of the relationship between connectivity and network dynamics allows us to directly design minimal, low-rank connectivity structures that implement specific computations. We focus on four tasks of increasing complexity, starting with basic binary discrimination and ending with context-dependent evidence integration (Mante et al., 2013). We find that the dynamical repertoire of the network increases quickly with the dimensionality of the connectivity structure, so that rank-two connectivity structures are already sufficient to implement complex, context-dependent tasks (Mante et al., 2013; Saez et al., 2015). For each task, we illustrate the relationship between connectivity, low-dimensional dynamics, and the performed computation. In particular, our framework naturally captures the ubiquitous observation that single-neuron responses are highly heterogeneous and mixed (Rigotti et al., 2013; Mante et al., 2013; Churchland and Shenoy, 2007; Machens et al., 2010), while the dimensionality of the dynamics underlying computations is low and increases with task complexity (Gao and Ganguli, 2015). Crucially, for each task, our framework produces experimentally testable predictions that directly relate connectivity, the dominant dimensions of the dynamics, and the computational features of individual neurons.

RESULTS

We studied a class of models that we call low-rank recurrent networks. In these networks, the connectivity matrix was given by a sum of an uncontrolled, random matrix and a structured, controlled matrix P. The structured matrix P was low-rank, i.e., it consisted only of a small number of independent rows and columns, and its entries were assumed to be weak (of order $1/N$, where N is the number of units in the network). We considered P moreover to be fixed and known, and uncorrelated with the random part $g\chi$, which was considered unknown except for its statistics (mean 0, variance g^2/N). As in classical models, the networks consisted of N firing rate units with a sigmoid input-output transfer function (Sompolinsky et al., 1988; Sussillo and Abbott, 2009):

$$\dot{x}_i(t) = -x_i(t) + \sum_{j=1}^{N} J_{ij}\phi(x_j(t)) + I_i, \qquad \text{(Equation 1)}$$

where $x_i(t)$ is the total input current to unit i, $J_{ij} = g\chi_{ij} + P_{ij}$ is the connectivity matrix, $\phi(x) = \tanh(x)$ is the current-to-rate transfer function, and I_i is the external, feedforward input to unit i.

To connect with the previous literature and introduce the methods that underlie our results, we start by describing the spontaneous dynamics $(I_i = 0)$ in a network with a unit-rank structure P. We then turn to the response to external inputs, the core of our results that we exploit to demonstrate how low-rank networks can implement four tasks of increasing complexity.

One-Dimensional Spontaneous Activity in Networks with Unit-Rank Structure

We started with the simplest possible type of low-dimensional connectivity, a matrix P with unit rank (Figure 1A). Such a matrix is specified by two N-dimensional vectors $m = \{m_i\}$ and $n = \{n_j\}$, which fully determine all its entries. Every column in this matrix is a multiple of the vector m, and every row is a multiple of the vector n, so that the individual entries are given by

$$P_{ij} = \frac{m_i n_j}{N}. \qquad \text{(Equation 2)}$$

We will call m and n, respectively, the right- and left-connectivity vectors (as they correspond to the right and left eigenvectors of the matrix P, see STAR Methods), and we consider them arbitrary but fixed and uncorrelated with the random part of the connectivity. As we will show, the spontaneous network dynamics can be directly understood from the geometrical arrangement of the vectors m and n.

In absence of structured connectivity, the dynamics are determined by the strength g of the random connectivity: for $g < 1$, the activity in absence of inputs decays to zero, while for $g > 1$ it displays strong, chaotic fluctuations (Sompolinsky et al., 1988). Our first aim was to understand how the interplay between the fixed,

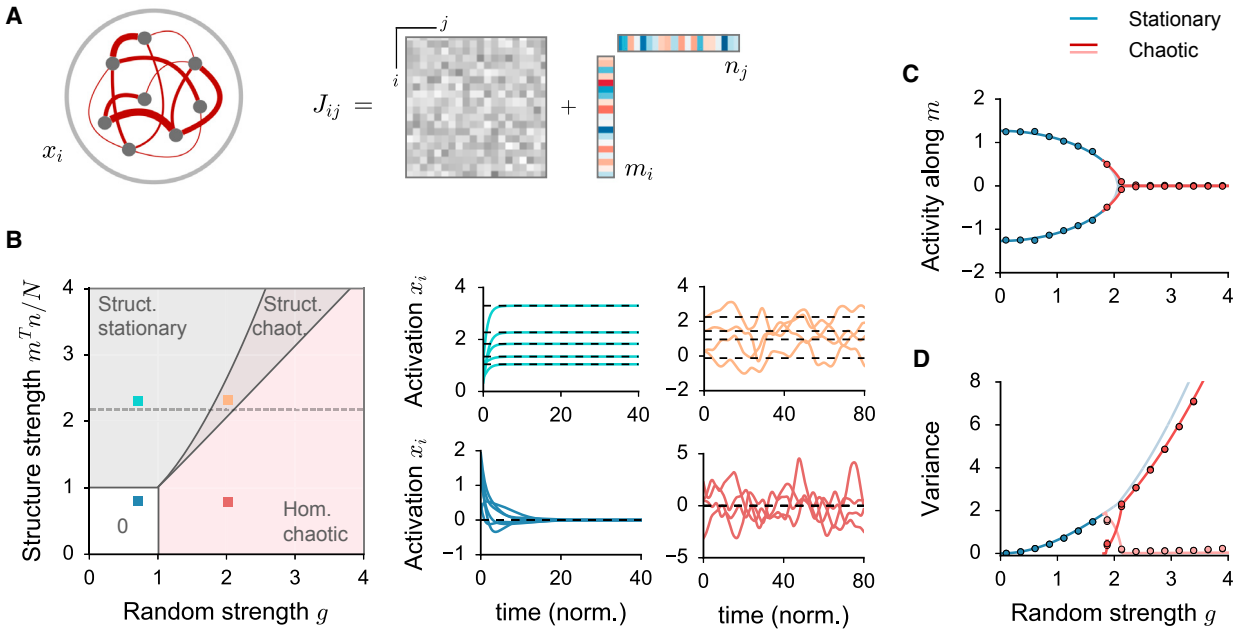

Figure 1. Spontaneous Activity in Random Networks with Unit-Rank Connectivity Structure

(A) The recurrent network model, whose connectivity matrix consists of the sum of a random (gray) and of a structured unit-rank (colored) component.

(B) Left: dynamical regimes of the network activity as function of the structure connectivity strength $m^T n/N$ and the random strength g. Gray areas: bistable activity; red: chaotic activity. Side panels: samples of dynamics from finite networks simulations (parameters indicated by colored dots in the phase diagram).

(C and D) Activity statistics as the random strength g is increased, and the structure strength is fixed to 2.2 (dashed line in B).

(C) Activity along the vector m, as quantified by κ (Equation 3). Blue (resp. red) lines: theoretical prediction for stationary (resp. chaotic) dynamics.

(D) Activity variance due to random connectivity. Blue and pink lines: static heterogeneity; red: temporal variance that quantifies chaotic activity. Dots: simulations of finite-size networks.

See STAR Methods for details.

low-rank part and the random part of the connectivity shapes the spontaneous activity in the network.

Our analysis of network dynamics relies on an effective, statistical description that can be mathematically derived if the network is large and the low-dimensional part of the connectivity is weak (i.e., if P_{ij} scales inversely with the number of units N in the network as in Equation 2). Under those assumptions, the activity of each unit can be described in terms of the mean and variance of the total input it receives. Dynamical equations for these quantities can be derived by extending the classical dynamical mean-field theory (Sompolinsky et al., 1988). This theory effectively leads to a low-dimensional description of network dynamics in terms of equations for a couple of macroscopic quantities. Full details of the analysis are provided in the STAR Methods; here, we focus only on the main results.

The central ingredient of the theory is an equation for the average equilibrium input μ_i to unit i:

$$\mu_i = \kappa m_i, \text{ where } \kappa = \frac{1}{N}\sum_{j=1}^{N} n_j [\phi_j]. \qquad \text{(Equation 3)}$$

The scalar quantity κ represents the overlap between the left-connectivity vector n and the N-dimensional vector $[\phi] = \{[\phi_j]\}$ that describes the mean firing activity of the network ($[\phi_j]$ is the firing rate of unit j averaged over different realizations of the random component of the connectivity, and depends implicitly on κ). The overlap κ therefore quantifies the degree of structure along the vector n in the activity of the network. If $\kappa > 0$, the equilibrium activity of each neuron is correlated with the corresponding component of the vector n, while $\kappa = 0$ implies no such structure is present. The overlap κ is the key macroscopic quantity describing the network dynamics, and our theory provides equations specifying its dependence on network parameters.

If one represents the network activity as a point in the $N-$dimensional state space where every dimension corresponds to the activity of a single unit, Equation 3 shows that the structured part of the connectivity induces a one-dimensional organization of the spontaneous activity along the vector m. This one-dimensional organization, however, emerges only if the overlap κ does not vanish. As the activity of the network is organized along the vector m, and κ quantifies the projection of the activity onto the vector n, non-vanishing values of κ require a non-vanishing overlap between vectors m and n. This overlap, given by $m^T n/N = \sum_j m_j n_j/N$, directly quantifies the strength of the structure in the connectivity. The connectivity structure strength $m^T n/N$ and the activity structure strength κ are therefore directly related, but in a highly non-linear manner. If the connectivity structure is weak, the network only exhibits homogeneous, unstructured activity corresponding to $\kappa = 0$ (Figure 1B, blue). If the connectivity structure is strong, structured

heterogeneous activity emerges ($\kappa > 0$), and the activity of the network at equilibrium is organized in one dimension along the vector m (Figures 1B, green, and 1C), while the random connectivity induces additional heterogeneity along the remaining $N-1$ directions. Note that because of the symmetry in the specific input-output function we use, when a heterogeneous equilibrium state exists, the configuration with the opposite sign is an equilibrium state too, so that the network activity is bistable (for more general asymmetric transfer functions, this bistability is still present, although the symmetry is lost, see Figure S7).

The random part of the connectivity disrupts the organization of the activity induced by the connectivity structure through two different effects. The first effect is that as the random strength g is increased, for any given realization of the random part of the connectivity, the total input to unit i will deviate more strongly from the expected mean μ_i (Figure 1D). As a consequence, the activity along the $N-1$ directions that are orthogonal to m increases, resulting in a noisy input to individual neurons that smoothens the gain of the non-linearity. This effectively leads to a reduction of the overall structure in the activity as quantified by κ (Figure 1C). A second, distinct effect is that increasing the random strength eventually leads to chaotic activity as in purely random networks. Depending on the strength of the structured connectivity, two different types of chaotic dynamics can emerge. If the disorder in the connectivity is much stronger than structure, the overlap κ is zero (Figure 1C). As a result, the mean activity of all units vanishes and the dynamics consist of unstructured, $N-$ dimensional temporal fluctuations (Figure 1D), as in the classical chaotic state of fully random networks (Figure 1B, red). In contrast, if the strengths of the random and structured connectivity are comparable, a structured type of chaotic activity emerges, in which $\kappa > 0$ so that the mean activity of different units is organized in one dimension along the direction m as shown by Equation 3, but the activity of different units now fluctuates in time (Figure 1B, orange). As for structured static activity, in this situation the system is bistable as states with opposite signs of κ always exist.

The phase diagram in Figure 1B summarizes the different types of spontaneous dynamics that can emerge as a function of the strength of structured and random components of the connectivity matrix. Altogether, the structured component of connectivity favors a one-dimensional organization of network activity, while the random component favors high-dimensional, chaotic fluctuations. Particularly interesting activity emerges when the structure and disorder are comparable, in which case the dynamics show one-dimensional structure combined with high-dimensional temporal fluctuations that can give rise to dynamics with very slow timescales (see Figure S6).

Two-Dimensional Activity in Response to an External Input

We now turn to the response to an external, feedforward input (Figure 2A). At equilibrium, the total average input to unit i is the sum of a recurrent input κm_i and the feedforward input I_i:

$$\mu_i = \kappa m_i + I_i, \text{ where } \kappa = \frac{1}{N}\sum_{j=1}^{N} n_j \left[\phi_j\right]. \qquad \text{(Equation 4)}$$

Transient, temporal dynamics close to this equilibrium are obtained by including temporal dependencies in κ and I_i (see STAR Methods; Equation 102).

Figure 2B illustrates the response of the network to a step input. The response of individual units is highly heterogeneous, different units showing increasing, decreasing, or multi-phasic responses. While every unit responds differently, the theory predicts that, at the level of the N-dimensional state space representing the activity of the whole population, the trajectory of the activity lies on average on the two-dimensional plane spanned by the right-connectivity vector m and the vector $I = \{I_i\}$ that corresponds to the pattern of external inputs (Figure 2B). Applying to the simulated activity a dimensionality reduction technique (see Cunningham and Yu [2014] for a recent review) such as principal-component analysis confirms that the two dominant dimensions of the activity indeed lie in the $m-I$ plane (Figure 2C), while the random part of connectivity leads to additional activity in the remaining $N-2$ directions that grows quickly with the strength of random connectivity g (see Figure S3). This approach therefore directly links the connectivity in the network to the emerging low-dimensional dynamics and shows that the dominant dimensions of activity are determined by a combination of feedforward inputs and connectivity (Wang et al., 2018).

The contribution of the connectivity vector m to the two-dimensional trajectory of activity is quantified by the overlap κ between the network activity $[\phi]$ and the left-connectivity vector n (Equation 4). If $\kappa = 0$, the activity trajectory is one dimensional and simply propagates the pattern of feedforward inputs. This is in particular the case for fully random networks. If $\kappa \neq 0$, the network response is instead a non-trivial two-dimensional combination of the input and connectivity structure patterns. In general, the value of κ, and therefore the organization of network activity, depends on the geometric arrangement of the input vector I with respect to the connectivity vectors m and n, as well as on the strength of the random component of the connectivity g.

As the neural activity lies predominantly in the $m-I$ plane, a non-vanishing κ, together with non-trivial two-dimensional activity, is obtained when the vector n has a non-zero component in the $m-I$ plane. Two qualitatively different input-output regimes can be distinguished. The first one is obtained when the connectivity vectors m and n are orthogonal to each other (Figure 2D, left and center). In that case, the overlap between them is zero, and the spontaneous activity in the network bears no sign of the underlying connectivity structure. Adding an external input can, however, reveal this connectivity structure and generate non-trivial two-dimensional activity if the input vector I has a non-zero overlap with the left-connectivity vector n. In such a situation, the vector n picks up the component of the activity along the feedforward input direction I. This leads to a non-zero overlap κ, which in turn implies that the network activity will have a component along the right-connectivity vector m. Increasing the external input along the direction of n will therefore progressively increase the response along m (Figure 2D, center), leading to a two-dimensional output.

A second, qualitatively different input-output regime is obtained when the connectivity vectors m and n have a strong enough overlap along a common direction (Figure 2D, right).

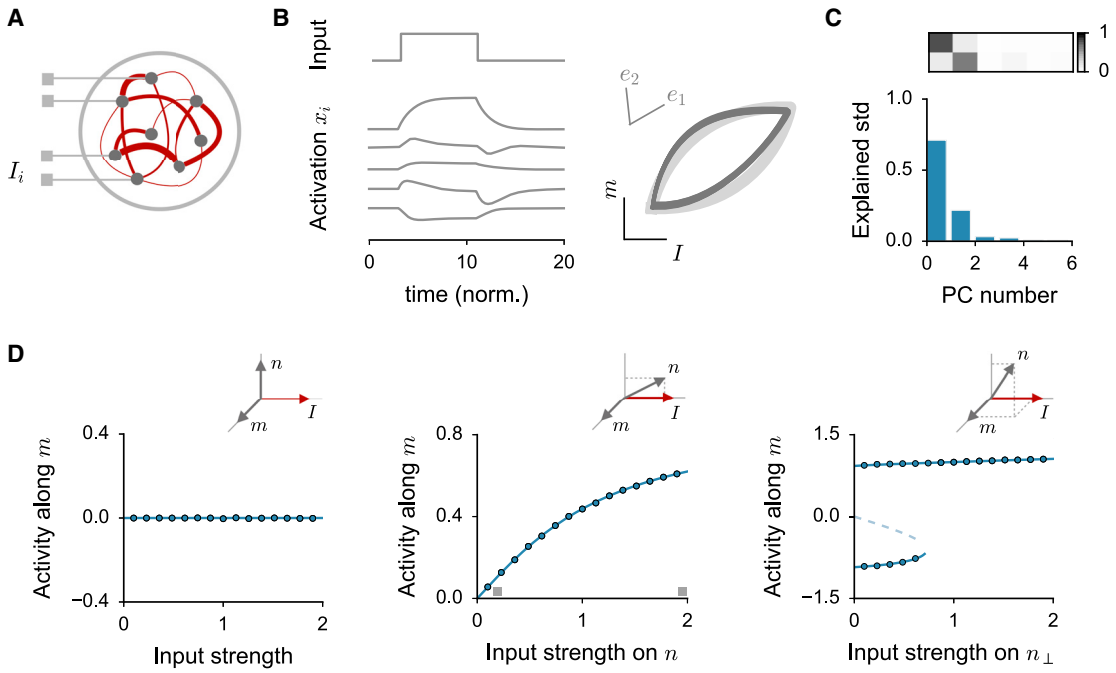

Figure 2. External Inputs Generate Two-Dimensional Activity in Random Networks with Unit-Rank Structure

(A) The pattern of external inputs can be represented by an N-dimensional vector $I = \{I_i\}$, where I_i is the input to unit i.

(B) Transient dynamics in response to a step input along I in a sample network of $N = 3500$ units. Left: activity traces for five units. Right: projections of the population trajectory onto the plane defined by the right-connectivity vector m and the input vector I. Light trace: theoretical prediction. Dark traces: simulations.

(C) Principal-component analysis (PCA) of the average activity trajectory. Bottom: fraction of SD explained by successive PCs. Top: correlation between PCs and the vectors m and I. The direction of the projections onto the $m - I$ plane of the two top PCs e_1 and e_2 are represented in (B). See also Figure S3.

(D) The activity κ along m is determined by the geometrical arrangement of the vector I and the connectivity vectors m and n. Three different cases are illustrated: (left) I, m, and n mutually orthogonal; (center) m and n mutually orthogonal, but I has a non-zero overlap with n; (right) m and n have non-zero overlap, leading to bistable activity in absence of inputs. Increasing the external input along n suppresses one of the two stable states. Continuous lines: theoretical predictions. Dots: simulations.

See STAR Methods for details.

As already shown in Figure 1, an overlap larger than unity between m and n induces bistable, structured spontaneous activity along the dimension m. Adding an external input along the vector n increases the activity along m but also eventually suppresses one of the bistable states. Large external inputs along the n direction therefore reliably set the network into a state in which the activity is a two-dimensional combination of the input direction and the connectivity direction m. This can lead to a strongly non-linear input-output transformation if the network was initially set in the state that lies on the opposite branch (Figure 2D, right).

An additional effect of an external input is that it generally tends to suppress chaotic activity present when the random part of connectivity is strong (Figures S3 and S4). This suppression occurs irrespectively of the specific geometrical configuration between the input I and connectivity vectors m and n and therefore independently of the two input-output regimes described above. Altogether, external inputs suppress both chaotic and bistable dynamics (Figure S4) and therefore always decrease the amount of variability in the dynamics (Churchland et al., 2010; Rajan et al., 2010).

In summary, external, feedforward inputs to a network with unit-rank connectivity structure in general lead to two-dimensional trajectories of activity. The elicited trajectory depends on the geometrical arrangement of the pattern of inputs with respect to the connectivity vectors m and n, which play different roles. The right-connectivity vector m determines the output pattern of network activity, while the left-connectivity vector n instead selects the inputs that give rise to outputs along m. An output structured along m can be obtained when n selects recurrent inputs (non-zero overlap between n and m) or when it selects external inputs (non-zero overlap between n and I).

Higher-Rank Structure Leads to a Rich Dynamical Repertoire

This far we focused on unit-rank connectivity structure, but our framework can be directly extended to higher-rank structure. A more general structured component of rank $r \ll N$ can be written as a superposition of r independent unit-rank terms

$$P_{ij} = \frac{m_i^{(1)} n_j^{(1)}}{N} + \ldots + \frac{m_i^{(r)} n_j^{(r)}}{N}, \qquad \text{(Equation 5)}$$

and is in principle characterized by $2r$ vectors $m^{(k)}$ and $n^{(k)}$. In such a network, the average dynamics lie in the $(r + 1)$-dimensional

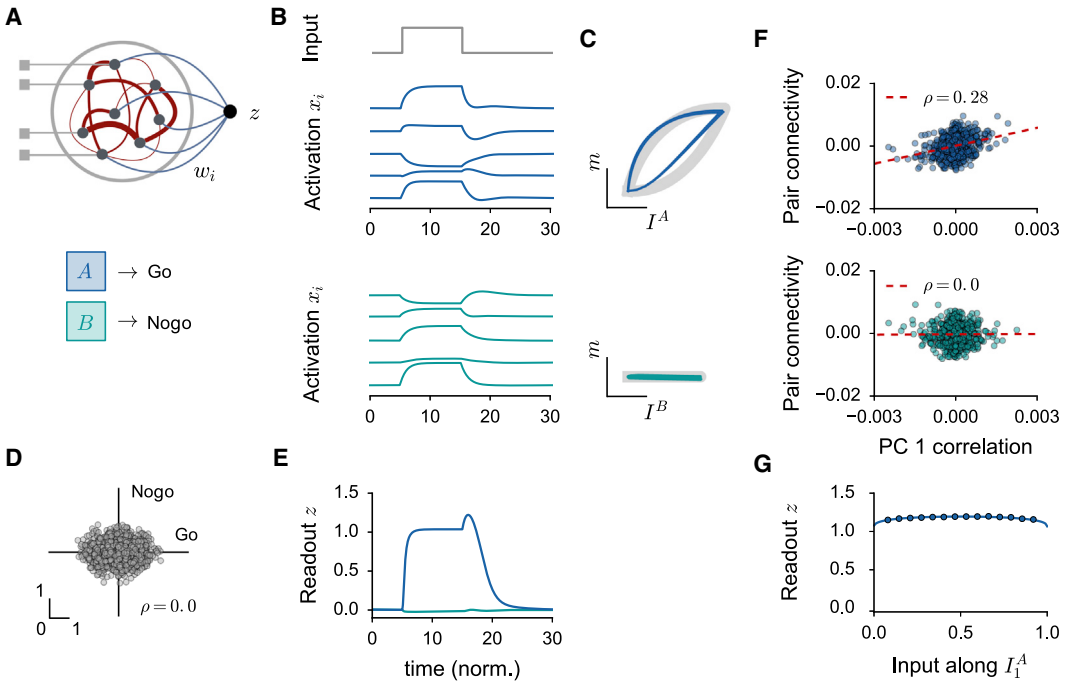

Figure 3. Implementing a Simple Go-Nogo Discrimination Task with a Unit-Rank Connectivity Structure

(A) A linear readout is added to the network, with randomly chosen weights w_i. The stimuli are represented by random input patterns I^A and I^B. The task consists in producing an output in response to stimulus *A*, but not *B*. The simplest unit-rank structure that implements the task is given by $m = w$ and $n = I^A$.

(B) Response of a sample network to the Go (blue) and Nogo (green) inputs. Activity traces for five units.

(C) Projections of the population trajectories onto the planes predicted to contain the dominant part of the dynamics. Gray: predicted trajectory. Colored traces: simulations.

(D) Linear regression coefficients for the Go and the Nogo stimuli. Every dot corresponds to a network unit.

(E) Readout dynamics for the Go (blue) and the Nogo (green) stimulus.

(F) Average connectivity strength as a function of the product between the coefficients of the first PC. Every dot corresponds to a pair of units.

(G) Generalization properties of the network. We select two Go stimuli I_1^A and I_2^A, and we set $n = I_1^A + I_2^A$. We build the input pattern as a normalized mixture of the two preferred patterns, and we gradually increase the component along I_1^A. Continuous lines: theoretical predictions. Dots: simulations.

See STAR Methods for details.

subspace spanned by the r right-connectivity vectors $m^{(k)}, k = 1, \ldots, r$ and the input vector I, while the left connectivity vectors $n^{(k)}$ select the inputs amplified along the corresponding dimension $m^{(k)}$. The details of the dynamics will in general depend on the geometrical arrangement of these $2r$ vectors among themselves and with respect to the input pattern. The number of possible configurations increases quickly with the structure rank, leading to a wide repertoire of dynamical states that includes continuous attractors (Figure S5) and sustained oscillatory activity (Figure S8). In the remainder of this manuscript, we will explore only the rank-two case.

Implementing a Simple Discrimination Task

Having developed an intuitive, geometric understanding of how a given unit-rank connectivity structure determines the low-dimensional dynamics in a network, we now reverse our approach to ask how a given computation can be implemented by choosing appropriately the structured part of the connectivity. We start with the computation underlying one of the most basic and most common behavioral tasks, Go-Nogo stimulus discrimination. In this task, an animal has to produce a specific motor output, e.g., press a lever or lick a spout, in response to a stimulus I^A (the Go stimulus), and ignore another stimuli I^B (Nogo stimuli). This computation can be implemented in a straightforward way in a recurrent network with a unit-rank connectivity structure. While such a simple computation does not in principle require a recurrent network, the implementation we describe here illustrates in a transparent manner the relationship between connectivity, dynamics, and computations in low-rank networks and leads to non-trivial and directly testable experimental predictions. It also provides the basic building block for more complex tasks, which we turn to in the next sections.

We model the sensory stimuli as random patterns of external inputs to the network, so that the two stimuli are represented by two fixed, randomly chosen N-dimensional vectors I^A and I^B. To model the motor response, we supplement the network with an output unit, which produces a linear readout $z(t) = \frac{1}{N}\sum_i w_i \varphi(x_i(t))$ of network activity (Figure 3A). The readout weights w_i are chosen randomly and form also a fixed N-dimensional vector w. The task of the network is to produce an output that is selective to the Go stimulus: the readout z at the end of

stimulus presentation needs to be non-zero for the input pattern I^A that corresponds to the Go stimulus, and zero for the other input I^B.

The two N-dimensional vectors m and n that generate the appropriate unit-rank connectivity structure to implement the task can be directly determined from our description of network dynamics. As shown in Equation 4 and Figure 2, the response of the network to the input pattern I is in general two-dimensional and lies in the plane spanned by the vectors m and I. The output unit will therefore produce a non-zero readout only if the readout vector w has a non-vanishing overlap with either m or I. As w is assumed to be uncorrelated, and therefore orthogonal, to all input patterns, this implies that the connectivity vector m needs to have a non-zero overlap with the readout vector w for the network to produce a non-trivial output. This output will depend on the amount of activity along m, quantified by the overlap κ. As shown in Figure 2, the overlap κ will be non-zero only if n has a non-vanishing overlap with the input pattern. Altogether, implementing the Go-Nogo task therefore requires that the right-connectivity vector m is correlated with the readout vector w, and that the left-connectivity vector n is correlated with the Go stimulus I^A.

Choosing $m = w$ and $n = I^A$ therefore provides the simplest unit-rank connectivity that implements the desired computation. Figure 3 illustrates the activity in the corresponding network. At the level of individual units, by construction both stimuli elicit large and heterogeneous responses (Figure 3B) that display mixed selectivity (Figure 3D). As predicted by the theory, the response to stimulus B is dominantly one-dimensional and organized along the input direction I^B, while the response to stimulus A is two-dimensional and lies in the plane defined by the right-connectivity vector m and the input direction I^A (Figure 3C). The readout from the network corresponds to the projection of the activity onto the m direction and is non-zero only in response to stimulus A (Figure 3E), so that the network indeed implements the desired Go-Nogo task. Our framework therefore allows us to directly link the connectivity, the low-dimensional dynamics, and the computation performed by the network and leads to two experimentally testable predictions. The first one is that performing a dimensionality reduction separately on responses to the two stimuli should lead to larger dimensionality of the trajectories in response to the Go stimulus. The second prediction is that for the Go stimulus, the dominant directions of activity depend on the recurrent connectivity in the network, while for the Nogo stimulus they do not. More specifically, for the activity elicited by the Go stimulus, the dominant principal components are combinations of the input vector I^A and right-connectivity vector m. Therefore, if two neurons have large principal-component weights, they are expected to also have large m weights and therefore stronger mutual connections than average (Figure 3F, top). In contrast, for the activity elicited by the Nogo stimulus, the dominant principal components are determined solely by the feedforward input, so that no correlation between dominant PC weights and recurrent connectivity is expected (Figure 3F, bottom). This prediction can in principle be directly tested in experiments analogous to Ko et al. (2011), where calcium imaging in behaving animals is combined with measurements of connectivity in a subset of recorded neurons. Note that in this setup very weak structured connectivity is sufficient to implement computations, so that the expected correlations may be weak if the random part of the connectivity is strong (see Figure S5).

The unit-rank connectivity structure forms the fundamental scaffold for the desired input-output transform. The random part of the connectivity adds variability around the target output and can induce additional chaotic fluctuations. Summing the activity of individual units through the readout unit, however, averages out this heterogeneity, so that the readout error decreases with network size as $1/\sqrt{N}$ (Figure S5). The present implementation is therefore robust to noise and has desirable computational properties in terms of generalization to novel stimuli. In particular, it can be extended in a straightforward way to the detection of a category of Go stimuli, rather than a single stimulus (Figure 3G).

Detection of a Noisy Stimulus

We now turn to a slightly more complex task: integration of a continuous, noisy stimulus. In contrast to the previous discrimination task, where the stimuli were completely different (i.e., orthogonal), here we consider a continuum of stimuli that differ only along the intensity of a single feature, such as the coherence of a random-dot kinetogram (Newsome et al., 1989). In a given stimulus presentation, this feature moreover fluctuates in time. We therefore represent each stimulus as $c(t)I$, where I is a fixed, randomly chosen input vector that encodes the relevant stimulus feature, and $c(t)$ is the amplitude of that feature. We consider a Go-Nogo version of this task, in which the network has to produce an output only if the average value of c is larger than a threshold (Figure 4A).

As for the basic discrimination task, the central requirements for a unit-rank network to implement this task are that the right-connectivity vector m is correlated with the readout vector w, and the left-connectivity vector n is correlated with the input pattern I. A key novel requirement in the present task is, however, that the response needs to be non-linear to produce the Go output when the strength of the input along I is larger than the threshold. As shown in Figure 2D, such a non-linearity can be obtained when the left- and right-connectivity vectors n and m have a strong enough overlap. We therefore add a shared component to m and n along a direction orthogonal to both w and I. In that setup, if the stimulus intensity c is low, the network will be in a bistable regime, in which the activity along the direction m can take two distinct values for the same input (Figure 2D, right). Assuming that the lower state represents a Nogo output, and that the network is initialized in this state at the beginning of the trial, increasing the stimulus intensity c above a threshold will lead to a sudden jump, and therefore a non-linear detection of the stimulus. Because the input amplitude fluctuates noisily in time, whether such a jump occurs depends on the integrated estimate of the stimulus intensity. The timescale over which this estimate is integrated is determined by the time constant of the effective exponential filter describing the network dynamics. In our unit-rank network, this time constant is set by the connectivity strength, i.e., the overlap between the left- and right-connectivity vectors m and n, which also determines the value of the threshold. Arbitrarily large timescales can be obtained by adjusting this overlap close to the bifurcation value, in which case the threshold becomes arbitrarily small (Figure 4F). In this section,

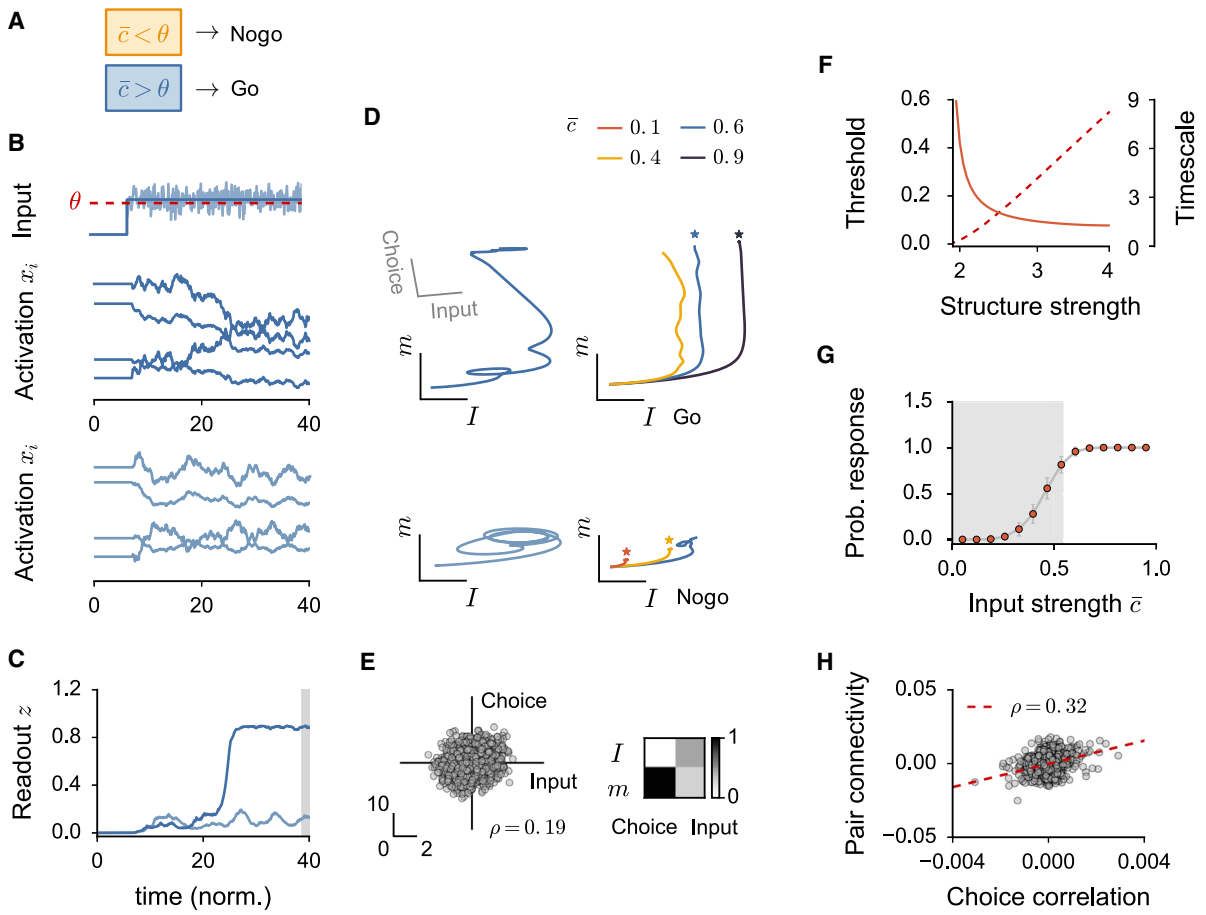

Figure 4. Implementing a Noisy Detection Task with a Unit-Rank Connectivity Structure

(A) The network is given a noisy input $c(t)$ along a fixed, random pattern of inputs I. The task consists in producing an output if the average input $\bar{c}$ is larger than a threshold θ.

(B) Dynamics in a sample network. Top: noisy input and threshold. Bottom: activity traces for four units and two different noise realizations in the stimulus, leading to a Go (dark blue) and a Nogo (light blue) output.

(C) Readout dynamics for the two stimuli.

(D) Projections of the population trajectory onto the plane defined by the right-connectivity vector m and the input vector I. Left: single-trial trajectories corresponding to (B). Right: trial-averaged trajectories, for Go (top) and Nogo (bottom) outputs, and different values of the mean input $\bar{c}$. Stars indicate correct responses.

(E) Left: linear regression coefficients for the input amplitude and the decision outcome. Every dot corresponds to a network unit. Right: correlation coefficients between the vectors m and I and the input and choice regression axes (see STAR Methods). Projection directions of the two input and choice regression axes onto the $m - I$ plane are shown in (D).

(F) Detection threshold (dashed) and timescale of the effective exponential filter (full line) for increasing values of the structure strength.

(G) Psychometric curve. The shaded area indicates the bistable region.

(H) Average connectivity strength as a function of the product of the linear regression coefficients for the choice variable. Every dot corresponds to a pair of network units.

See STAR Methods for details.

we fix the structure strength so that the threshold is set to 0.5, which corresponds to an integration timescale of the order of the time constant of individual units.

Figure 4 illustrates the activity in an example implementation of this network. In a given trial, as the stimulus is noisy, the activity of the individual units fluctuates strongly (Figure 4B). Our theory predicts that the population trajectory on average lies in the plane defined by the connectivity vector m and the input pattern I (Figure 4D). Activity along the m direction is picked up by the readout, and its value at the end of stimulus presentation determines the output (Figure 4C). Because of the bistable dynamics in the network, whether the m direction is explored, and an output produced, depends on the specific noisy realization of the stimulus. Stimuli with an identical average strength can therefore lead to either two-dimensional trajectories of activity and Go responses or one-dimensional trajectories of activity corresponding to Nogo responses (Figure 4D). The probability of generating an output as function of stimulus strength follows a sigmoidal psychometric curve that reflects the underlying bistability (Figure 4G). Note that the

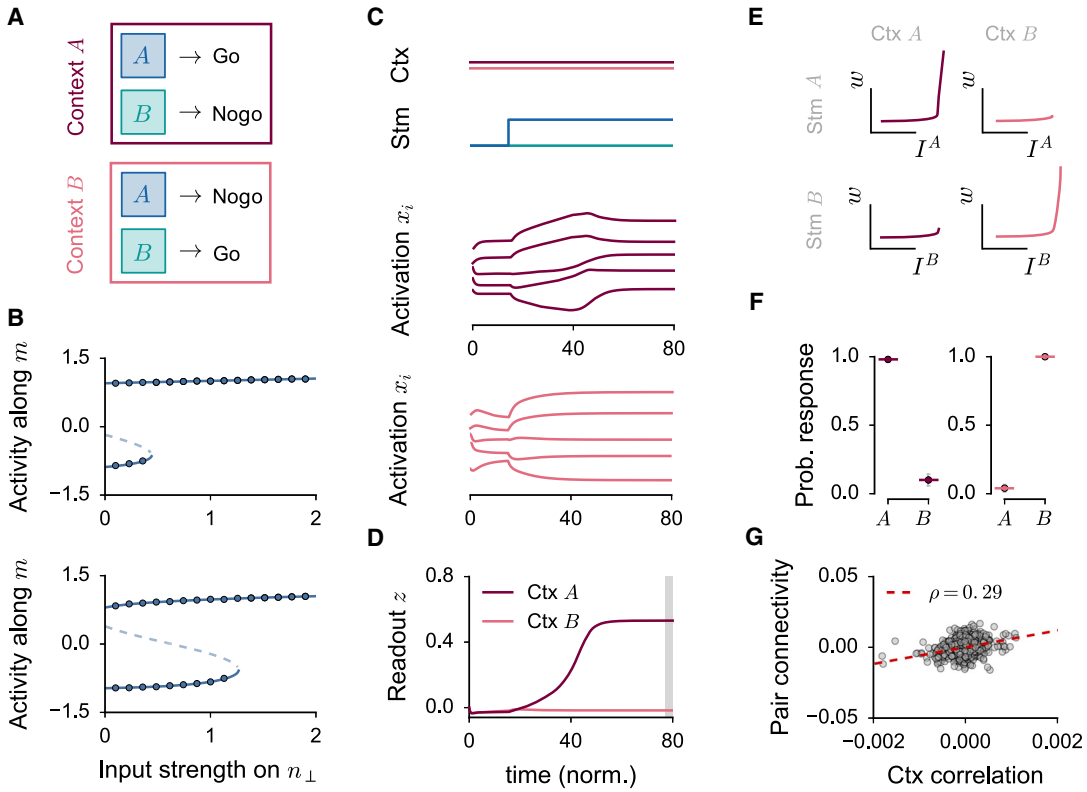

Figure 5. Implementing a Context-Dependent Go-Nogo Discrimination Task with a Rank-Two Connectivity Structure

(A) As in Figure 3, two stimuli *A* and *B* are presented to the network. The task consists in producing an output in response to the Go stimulus, which is determined by the contextual cue (*A* in context *A*, *B* in context *B*), modeled as inputs along random directions I_{ctxA} and I_{ctxB}.

(B) Inputs along the overlap direction between the left- and the right-connectivity vectors modulate the response threshold of the network (see also Figure S5).

(C) Dynamics in a sample network in response to the stimulus *A*. Top: stimulus and contextual input. Bottom: activity for five units in context *A* (crimson) and *B* (pink).

(D) Readout dynamics in the two contexts.

(E) Projections of the average population trajectories onto the planes spanned by vectors *w*, I^A and I^B.

(F) Network performance in the two contexts.

(G) Average connectivity strength between pairs of units as a function of the product between the regression coefficients for context. Every dot corresponds to a pair of network units.

See STAR Methods for details.

bistability is not clearly apparent on the level of individual units. In particular, the activity of individual units is always far from saturation, as their inputs are distributed along a zero-centered Gaussian (Equation 4).

The responses of individual units are strongly heterogeneous and exhibit mixed selectivity to stimulus strength and output choice (Figure 4E). A popular manner to interpret such activity at the population level is a targeted dimensional reduction approach, in which input and choice dimensions are determined through regression analyses (Mante et al., 2013). As expected from our theoretical analysis, the two dimensions obtained through regression are closely related to *m* and *I*; in particular, the choice dimension is highly correlated with the right-connectivity vector *m* (Figure 4E). As a result, the plane in which network activity dominantly lies corresponds to the plane defined by the choice and the input dimensions (Figure 4D). Our framework therefore directly links recurrent connectivity and effective output choice direction through the low-dimensional dynamics. A resulting experimentally testable prediction is that neurons with strong choice regressors have stronger mutual connections (Figure 4H).

A Context-Dependent Discrimination Task

We next consider a context-dependent discrimination task, in which the relevant response to a stimulus depends on an additional, explicit contextual cue. Specifically, we focus on the task studied in Saez et al. (2015) where in one context (referred to as context *A*), the stimulus *A* requires a Go output, and the stimulus *B* a Nogo, while in the other context (referred to as context *B*), the associations are reversed (Figure 5A). This task is a direct extension of the basic binary discrimination task introduced in Figure 3; yet it is significantly more complex as it represents a hallmark of cognitive flexibility: a non-linearly separable, *XOR*-like computation that a single-layer feedforward network cannot solve (Rigotti et al., 2010; Fusi et al., 2016). We will show that this task can be implemented in a rank-two recurrent

network that is a direct extension of the unit-rank network used for the discrimination task in Figure 4.

This context-dependent task can be seen as a combination of two basic, opposite Go-Nogo discriminations, each of which can be independently implemented by a unit-rank structure with the right-connectivity vector m correlated to the readout, and the left-connectivity vector correlated to the Go input (I^A for context A, I^B for context B). Combining two such unit-rank structures, with left-connectivity vectors $n^{(1)}$ and $n^{(2)}$ correlated respectively with I^A and I^B, leads to a rank-two connectivity structure that serves as a scaffold for the present task. The cues for context A and B are represented by additional inputs along random vectors I_{ctxA} and I_{ctxB}, presented for the full length of the trial (Remington et al., 2018) (Figure 5C). These inputs are the only contextual information incorporated in the network. In particular, the readout vector w is fixed and independent of the context (Mante et al., 2013). Crucially, since the readout w needs to produce an output for both input stimuli, both right-connectivity vectors $m^{(1)}$ and $m^{(2)}$ need to be correlated with it.

The key requirement for implementing context-dependent discrimination is that each contextual input effectively switches off the irrelevant association. To implement this requirement, we rely on the same non-linearity as for the noisy discrimination task, based on the overlap between the left- and right-connectivity vectors (Figure 2D). We however exploit an additional property, which is that the threshold of the non-linearity (i.e., the position of the transition from a bistable to a mono-stable region in Figure 2D) can be controlled by an additional modulatory input along the overlap direction between m and n (Figures 5B and S4). Such a modulatory input acts as an effective offset for the bistability at the macroscopic, population level (see Equation 153 in STAR Methods). A stimulus of a given strength (e.g., unit strength in Figure 5B) may therefore induce a transition from the lower to the upper state (Figure 5B, top), or no transition (Figure 5B bottom) depending on the strength of the modulatory input that sets the threshold value. While in the noisy discrimination task, the overlap between m and n was chosen in an arbitrary direction, in the present setting, we take the overlaps between each pair of left- and right-connectivity vectors to lie along the direction of the corresponding contextual input (i.e., $m^{(1)}$ and $n^{(1)}$ overlap along I_{ctxA}, $m^{(2)}$ and $n^{(2)}$ along I_{ctxB}), so that contextual inputs directly modulate the threshold of the non-linearity. The final rank-two setup is described in detail in the STAR Methods.

Figure 5 illustrates the activity in an example of the resulting network implementation. The contextual cue is present from the very beginning of the trial and effectively sets the network in a context-dependent initial state (Figure 5C) that corresponds to the lower of the two bistable states. The low-dimensional response of the network to the following stimulus is determined by this initial state and the sustained contextual input. If the cue for context A is present, stimulus A leads to the crossing of the non-linearity, a transition from the lower to the upper state, and therefore a two-dimensional response in the plane determined by I^A and w (Figure 5E, top left), generating a Go output (Figure 5D). In contrast, if the cue for context B is present, the threshold of the underlying non-linearity is increased in the direction of input I^A (Figure 5B, bottom), so that the presentation of stimulus A does not induce a transition between the lower and upper states but leads only to a one-dimensional trajectory orthogonal to the readout, and therefore a Nogo response (Figure 5E, top right). The situation is totally symmetric in response to stimulus B (Figure 5E, bottom), so that contextual cues fully reverse the stimulus-response associations (Figure 5F). Overall, this context-dependent discrimination relies on strongly non-linear interactions between the stimulus and contextual inputs, that on the connectivity level are implemented by overlaps between the connectivity vectors along the contextual inputs. A central, experimentally testable prediction of our framework is therefore that, if a network is implementing this computation, units with strong contextual selectivity have on average stronger mutual connections (Figure 5G).

A Context-Dependent Evidence Integration Task

We finally examine a task inspired by Mante et al. (2013) that combines context-dependent output and fluctuating, noisy inputs. The stimuli now consist of superpositions of two different features A and B, and the strengths of both features fluctuate in time during a given trial. In Mante et al. (2013), the stimuli were random dot kinetograms, and the features A and B corresponded to the direction of motion and color of these stimuli. The task consists in classifying the stimuli according to one of those features, the relevant one being indicated by an explicit contextual cue (Figure 6A).

We implemented a Go-Nogo version of the task, in which the output is required to be non-zero when the relevant feature is stronger than a prescribed threshold (arbitrarily set to 0.5). The present task is therefore a direct combination of the detection task introduced in Figure 4 and the context-dependent discrimination task of Figure 5, but the individual stimuli are now two dimensional, as they consist of two independently varied features A and B. In this task, a significant additional difficulty is that on every trial the irrelevant feature needs to be ignored, even if it is stronger than the relevant feature (e.g., color coherence stronger than motion coherence on a motion-context trial).

This context-dependent evidence integration task can be implemented with exactly the same rank-two configuration as the basic context-dependent discrimination in Figure 5, with contextual gating relying on the same non-linear mechanism as in Figure 5B. The contextual cue is presented throughout the trial (Figure 6B) and determines which of the features of the two-dimensional stimulus leads to non-linear dynamics along the direction of connectivity vectors $m^{(1)}$ and $m^{(2)}$ (Figure 6D). These directions share a common component along the readout vector w, and the readout unit picks up the activity along that dimension. As a consequence, depending on the contextual cue, the same stimulus can lead to opposite outputs (Figure 6C). Altogether, in context A, the output is independent of the values of feature B, and conversely in context B (Figure 6E). The output therefore behaves as if it were based on two orthogonal readout directions, yet the readout direction is unique and fixed, and the output relies instead on a context-dependent selection of the relevant input feature (Mante et al., 2013).

An important additional requirement in the present task with respect to the basic context-dependent integration is that the network needs to perform temporal integration to average out temporal fluctuations in the stimulus. As illustrated in Figures

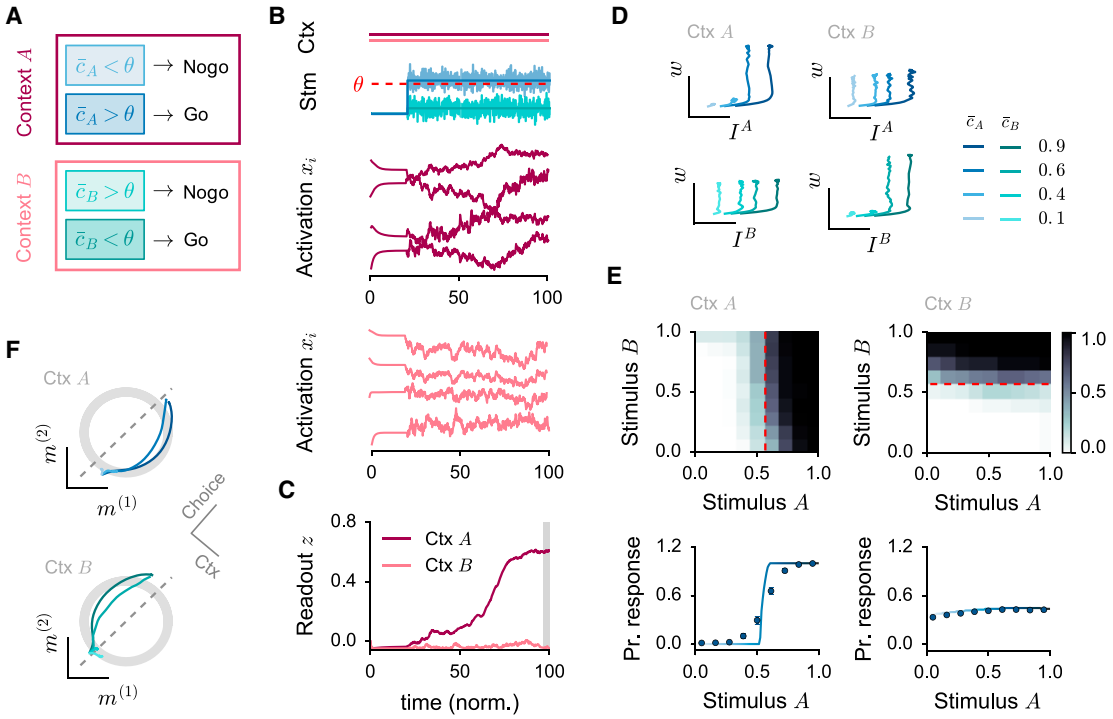

Figure 6. Implementing a Context-Dependent Evidence Accumulation Task Using Rank-Two Connectivity Structure

(A) The stimuli consist of a superposition of two features c_A and c_B, which fluctuate in time around mean values $\bar{c}_A$ and $\bar{c}_B$. In every trial, a pair of contextual inputs determines the relevant input feature. The task consists in producing an output if the average strength of the relevant feature is larger than a threshold.

(B) Dynamics in a sample network. Top: stimulus and contextual inputs. Bottom: activity of four units in contexts A (crimson) and B (pink).

(C) Readout dynamics in the two contexts.

(D) Average population trajectories projected onto the planes spanned by vectors w, I^A and I^B. Blue (resp. green) trajectories have been sorted according to the value of the strength of stimulus A (resp. B), and averaged across stimulus B (resp. A).

(E) Network performance. Top row: probability of response as function of input strengths $\bar{c}_A$ and $\bar{c}_B$ (simulated data). Bottom: probability of response averaged over $\bar{c}_B$. Continuous line: theoretical prediction; dots: simulations.

(F) Projection of the population activity onto the plane defined by the orthogonal components of the vectors m_A and m_B and comparison with the underlying circular attractor (see STAR Methods). Trajectories are sorted by the strength of the relevant stimulus and averaged across the non-relevant one. The direction of the projections of the regression axes for choice and context are indicated in gray.

See STAR Methods for details.

6B and 6C, the network dynamics in response to stimuli indeed exhibit a slow timescale and progressively integrate the input. Strikingly, such slow dynamics do not require additional constraints on network connectivity; they are a direct consequence of the rank-two connectivity structure used for contextual gating (in fact the dynamics are already slow in the basic contextual discrimination task, see Figures 5C and 5D). More specifically, the symmetry between the two contexts implies that two sets of left- and right-connectivity vectors have identical overlaps (i.e. $m^{(1)T}n^{(1)} = m^{(2)T}n^{(2)}$). Without further constraints on the connectivity, such a symmetric configuration leads to an emergence of a continuous line attractor, with the shape of a two-dimensional ring in the plane defined by $m^{(1)}$ and $m^{(2)}$ (see STAR Methods and Figure S5). In the implementation of the present task, on top of symmetric overlaps, the four connectivity vectors include a common direction along the readout vector. This additional constraint eliminates the ring attractor and stabilizes only two equilibrium states that correspond to Go and Nogo outputs. Yet, the ring attractor is close in parameter space, and this proximity induces a slow manifold in the dynamics, so that the trajectories leading to a Go output slowly evolve along two different sides of the underlying ring depending on the context (Figure 6F). As a result, the two directions in the plane $m^{(1)} - m^{(2)}$ correspond to choice and context axis as found by regression analysis (Figure 6F). A similar mechanism for context-dependent evidence integration based on a line attractor was previously identified by reverse-engineering a trained recurrent network (Mante et al., 2013). Whether the underlying dynamical structure was a ring as in our case or two line attractors for the two contexts depended on the details of the network training protocol (V. Mante, unpublished data). Here, we show that such a mechanism based on a ring attractor can be implemented in a minimal network with rank-two connectivity structure, but other solutions can certainly be found. Note that this rank-two network can also serve as an alternative implementation for context-independent evidence integration in which the integration timescale and the threshold value are fully independent in contrast to the unit-rank implementation (Figure 4).

DISCUSSION

Motivated by the observation that a variety of approaches for implementing computations in recurrent networks rely on a common type of connectivity structure, we studied a class of models in which the connectivity matrix consists of a sum of a fixed, low-rank term and a random part. Our central result is that the low-rank connectivity structure induces low-dimensional dynamics in the network, a hallmark of population activity recorded in behaving animals (Gao and Ganguli, 2015). While low-dimensional activity is usually detected numerically using dimensional-reduction techniques (Cunningham and Yu, 2014), we showed that a mean-field theory allows us to directly predict the low-dimensional dynamics based on the connectivity and input structure. This approach led us to a simple, geometrical understanding of the relationship between connectivity and dynamics and enabled us to design minimal-connectivity implementations of specific computations. In particular, we found that the dynamical repertoire of the network increases quickly with the rank of the connectivity structure, so that rank-two networks can already implement a variety of computations. In this study, we have not explicitly considered structures with rank higher than two, but our theoretical framework is in principle valid for arbitrary rank $r \ll N$, where N is the size of the network.

While other works have examined dynamics in networks with a mixture of structured and random connectivity (e.g., Roudi and Latham [2007]; Ahmadian et al. [2015]), the most classical approach for implementing computations in recurrent networks has been to endow them with a clustered (Wang, 2002; Amit and Brunel, 1997; Litwin-Kumar and Doiron, 2012) or distance-dependent connectivity (Ben-Yishai et al., 1995). Such networks inherently display low-dimensional dynamics similar to our framework (Doiron and Litwin-Kumar, 2014; Williamson et al., 2016), as clustered connectivity is in fact a special case of low-rank connectivity. Clustered connectivity, however, is highly ordered: each neuron belongs to a single cluster and therefore is selective to a single task feature (e.g., a given stimulus, or a given output). Neurons in clustered networks are therefore highly specialized and display pure selectivity (Rigotti et al., 2013). Here, instead, we have considered random low-rank structures, which generate activity organized along heterogeneous directions in state space. As a consequence, stimuli and outputs are represented in a random, highly distributed manner and individual neurons are typically responsive to several stimuli, outputs, or combinations of the two. Such mixed selectivity is a ubiquitous property of cortical neurons (Rigotti et al., 2013; Mante et al., 2013; Churchland and Shenoy, 2007) and confers additional computational properties to our networks (Kanerva, 2009). In particular, it allowed us to easily extend to a context-dependent situation (Mante et al., 2013; Saez et al., 2015), a network implementation of a basic discrimination task. This is typically difficult to do in clustered, purely selective networks (Rigotti et al., 2010).

The type of connectivity used in our study is closely related to the classical framework of Hopfield networks (Hopfield, 1982; Amit et al., 1985). The aim of Hopfield networks is to store in memory specific patterns of activity by creating for each pattern a corresponding fixed point in the network dynamics. This is achieved by adding a unit-rank term for each item, and one approach for investigating the capacity of such a setup has relied on the mean-field theory of a network with a connectivity that consists of a sum of a rank-one term and a random matrix (Tirozzi and Tsodyks, 1991; Shiino and Fukai, 1993; Roudi and Latham, 2007). While this approach is clearly close to the one adopted in the present study, there are important differences. Within Hopfield networks, the unit-rank terms are symmetric, so that the corresponding left- and right-connectivity vectors are identical for each pattern. Moreover, the unit-rank terms that correspond to different patterns are generally uncorrelated. In contrast, here we have considered the more general case where the left- and right-eigenvectors are different and potentially correlated between different rank-one terms. Most importantly, our main focus was on responses to external inputs and input-output computations, rather than memorizing items. In particular, we showed that left- and right-connectivity vectors play different roles with respect to processing inputs, with the left-connectivity vector implementing input- selection, and the right-connectivity vector determining the output of the network.

Our study is also directly related to echo-state networks (ESNs) (Jaeger and Haas, 2004) and FORCE learning (Sussillo and Abbott, 2009). In those frameworks, randomly connected recurrent networks are trained to produce specified outputs using a feedback loop from a readout unit to the network, which is mathematically equivalent to adding a rank-one term to the random connectivity matrix (Maass et al., 2007). In their most basic implementation, both ESN and FORCE learning train only the readout weights. The training is performed for a fixed, specified realization of the random connectivity, so that the final rank-one structure is correlated with the random part of the connectivity and may be strong with respect to it. In contrast, the results presented here rely on the assumption that the low-rank structure is weak and independent from the random part. Although ESN and FORCE networks do not necessarily fulfill this assumption, in ongoing work we found that our approach describes well networks trained using ESN or FORCE to produce a constant output (Rivkind and Barak, 2017). Note that in our framework, the computations rely solely on the structured part of the connectivity, but ongoing work suggests that the random part of the connectivity may play an important role during training.

The specific network model used here is identical to most studies based on trained recurrent networks (Sussillo and Abbott, 2009; Mante et al., 2013; Sussillo, 2014). It is highly simplified and lacks many biophysical constraints, the most basic ones being positive firing rates, the segregation between excitation and inhibition and interactions through spikes. Recent works have investigated extensions of the abstract model used here to networks with biophysical constraints (Ostojic, 2014; Kadmon and Sompolinsky, 2015; Harish and Hansel, 2015; Mastrogiuseppe and Ostojic, 2017; Thalmeier et al., 2016). Additional work will be needed to implement the present framework in networks of spiking neurons.

Our results imply novel, directly testable experimental predictions relating connectivity, low-dimensional dynamics and computational properties of individual neurons. Our main result is that the dominant components of low-dimensional dynamics are a combination of feedforward input patterns, and vectors

specifying the low-rank recurrent connectivity (Figure 2C). A direct implication is that, if the low-dimensional dynamics in the network are generated by low-rank recurrent connectivity, two neurons that have large loadings in the dominant principal components will tend to have mutual connections stronger than average (Figure 3F, top). In contrast, if the low-dimensional dynamics are not generated by recurrent interactions but instead are driven by feedforward inputs alone, no correlation between principal components and connectivity is expected (Figure 3F, bottom). Since the low-dimensional dynamics based on recurrent connectivity form the scaffold for computations in our model, this basic prediction can be extended to various task-dependent properties of individual neurons. For instance, if the recurrent connectivity implements evidence integration, two units with strong choice regressors are predicted to have mutual connections stronger than average (Figure 4H). Analogously, if recurrent connections implement context-dependent associations, two units with strong context regressors are expected to share connections stronger than average (Figure 5G). Such predictions can in principle be directly tested in experiments that combine calcium imaging of neural activity in behaving animals with measurements of connectivity between a subset of recorded neurons (Ko et al., 2011). It should be noted, however, that very weak structured connectivity is sufficient to implement computations, so that the expected correlations between connectivity and various selectivity indices may be weak.

The class of recurrent networks we considered here is based on connectivity matrices that consist of an explicit sum of a low-rank and a random part. While this may seem as a limited class of models, in fact, any arbitrary matrix can be approximated with a low-rank one, e.g., by keeping a small number of dominant singular values and singular vectors (Markovsky, 2012)—this is the basic principle underlying dimensionality reduction. A recurrent network with any arbitrary connectivity matrix can therefore in principle be approximated by a low-rank recurrent network. From this point of view, our theory suggests a simple conjecture: the low-dimensional structure in connectivity determines low-dimensional dynamics and computational properties of recurrent networks. While more work is needed to establish under which precise conditions a low-rank network provides a good computational approximation of a full recurrent network, this conjecture provides a simple and practically useful working hypothesis for reverse-engineering trained neural networks (Sussillo and Barak, 2013), and relating connectivity, dynamics, and computations in neural recordings.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.07.003.

ACKNOWLEDGMENTS

We are grateful to Alexis Dubreuil, Vincent Hakim, and Kishore Kuchibhotla for discussions and feedback on the manuscript. This work was funded by the Programme Emergences of City of Paris, Agence Nationale de la Recherche grants ANR-16-CE37-0016-01 and ANR-17-ERC2-0005-01, and the program "Investissements d'Avenir" launched by the French Government and implemented by the ANR, with the references ANR-10-LABX-0087 IEC and ANR-11-IDEX-0001-02 PSL* Research University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

F.M. and S.O. designed the study and wrote the manuscript. F.M. performed model analyses and simulations.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 4, 2017
Revised: April 27, 2018
Accepted: July 2, 2018
Published: July 26, 2018

REFERENCES

Ahmadian, Y., Fumarola, F., and Miller, K.D. (2015). Properties of networks with partially structured and partially random connectivity. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *91*, 012820.

Aljadeff, J., Stern, M., and Sharpee, T. (2015b). Transition to chaos in random networks with cell-type-specific connectivity. Phys. Rev. Lett. *114*, 088101.

Amit, D.J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cereb. Cortex *7*, 237–252.

Amit, D.J., Gutfreund, H., and Sompolinsky, H. (1985). Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys. Rev. Lett. *55*, 1530–1533.

Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. Curr. Opin. Neurobiol. *46*, 1–6.

Ben-Yishai, R., Bar-Or, R.L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. USA *92*, 3844–3848.

Boerlin, M., Machens, C.K., and Denève, S. (2013). Predictive coding of dynamical variables in balanced spiking networks. PLoS Comput. Biol. *9*, e1003258.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J. Comput. Neurosci. *8*, 183–208.

Churchland, M.M., and Shenoy, K.V. (2007). Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. J. Neurophysiol. *97*, 4235–4257.

Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., et al. (2010). Stimulus onset quenches neural variability: A widespread cortical phenomenon. Nat. Neurosci. *13*, 369–378.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.

Doiron, B., and Litwin-Kumar, A. (2014). Balanced neural architecture and the idling brain. Front. Comput. Neurosci. *8*, 56.

Eliasmith, C., and Anderson, C. (2004). Neural Engineering - Computation, Representation, and Dynamics in Neurobiological Systems (MIT Press).

Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. Curr. Opin. Neurobiol. *37*, 66–74.

Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. Curr. Opin. Neurobiol. *32*, 148–155.

Girko, V.L. (1985). Circular law. Theory Probab. Appl. *29*, 694–706.

Harish, O., and Hansel, D. (2015). Asynchronous rate chaos in spiking neuronal circuits. PLoS Comput. Biol. *11*, e1004266.

Harris, K.D., and Mrsic-Flogel, T.D. (2013). Cortical connectivity and sensory coding. Nature *503*, 51–58.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA *79*, 2554–2558.

Jaeger, H., and Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science *304*, 78–80.

Kadmon, J., and Sompolinsky, H. (2015). Transition to chaos in random neuronal networks. Phys. Rev. X *5*, 041030.

Kanerva, P. (2009). Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. Cognit. Comput. *1*, 139–159.

Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87–91.

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat. Neurosci. *15*, 1498–1505.

Maass, W., Joshi, P., and Sontag, E.D. (2007). Computational aspects of feedback in neural circuits. PLoS Comput. Biol. *3*, e165.

Machens, C.K., Romo, R., and Brody, C.D. (2010). Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. J. Neurosci. *30*, 350–360.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Markovsky, I. (2012). Low Rank Approximation - Algorithms, Implementations, Applications (Springer).

Martens, J., and Sutskever, I. (2011). Learning recurrent neural networks with hessian-free optimization. In ICML'11 Proceedings of the 28th International Conference on International Conference on Machine Learning. (ICML), pp. 1033–1040.

Mastrogiuseppe, F., and Ostojic, S. (2017). Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. PLoS Comput. Biol. *13*, e1005498.

Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. Nature *341*, 52–54.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. Nat. Neurosci. *17*, 594–600.

Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In ICML'13 Proceedings of the 30th International Conference on International Conference on Machine Learning (ICML), pp. III–1310–III–1318.

Rajan, K., and Abbott, L.F. (2006). Eigenvalue spectra of random matrices for neural networks. Phys. Rev. Lett. *97*, 188104.

Rajan, K., Abbott, L.F., and Sompolinsky, H. (2010). Stimulus-dependent suppression of chaos in recurrent neural networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *82*, 011903.

Remington, E.D., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. Neuron *98*, 1005–1019.e5.

Rigotti, M., Ben Dayan Rubin, D., Wang, X.-J., and Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: The importance of the diversity of neural responses. Front. Comput. Neurosci. *4*, 24.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590.

Rivkind, A., and Barak, O. (2017). Local dynamics in trained recurrent neural networks. Phys. Rev. Lett. *118*, 258101.

Roudi, Y., and Latham, P.E. (2007). A balanced memory network. PLoS Comput. Biol. *3*, 1679–1700.

Saez, A., Rigotti, M., Ostojic, S., Fusi, S., and Salzman, C.D. (2015). Abstract context representations in primate amygdala and prefrontal cortex. Neuron *87*, 869–881.

Shadlen, M.N., and Newsome, W.T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. J. Neurosci. *18*, 3870–3896.

Shiino, M., and Fukai, T. (1993). Self-consistent signal-to-noise analysis of the statistical behavior of analog neural networks and enhancement of the storage capacity. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics *48*, 867–897.

Sompolinsky, H., Crisanti, A., and Sommers, H.J. (1988). Chaos in random neural networks. Phys. Rev. Lett. *61*, 259–262.

Sussillo, D. (2014). Neural circuits as computational dynamical systems. Curr. Opin. Neurobiol. *25*, 156–163.

Sussillo, D., and Abbott, L.F. (2009). Generating coherent patterns of activity from chaotic neural networks. Neuron *63*, 544–557.

Sussillo, D., and Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput. *25*, 626–649.

Tao, T. (2013). Outliers in the spectrum of iid matrices with bounded rank perturbations. Probab. Theory Relat. Fields *155*, 231–263.

Thalmeier, D., Uhlmann, M., Kappen, H.J., and Memmesheimer, R.M. (2016). Learning universal computations with spikes. PLoS Comput. Biol. *12*, e1004895.

Tirozzi, B., and Tsodyks, M. (1991). Chaos in highly diluted neural networks. EPL *14*, 727.

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science *274*, 1724–1726.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron *36*, 955–968.

Wang, J., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. Nat. Neurosci. *21*, 102–110.

Williamson, R.C., Cowley, B.R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M.A., and Yu, B.M. (2016). Scaling properties of dimensionality reduction for neural populations and network models. PLoS Comput. Biol. *12*, e1005141.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Algorithms for solving Dynamical Mean-Field equations	this paper	https://github.com/fmastrogiuseppe/lowrank/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further requests for resources should be directed to and will be fulfilled by the Lead Contact, Srdjan Ostojic (srdjan.ostojic@ens.fr).

METHOD DETAILS

The network model

We study large recurrent networks of rate units. Every unit in the network is characterized by a continuous variable $x_i(t)$, commonly interpreted as the total input current. More generically, we also refer to $x_i(t)$ as the *activation* variable. The output of each unit is a non-linear function of its inputs modeled as a sigmoidal function $\phi(x)$. In line with previous works (Sompolinsky et al., 1988; Sussillo and Abbott, 2009; Rivkind and Barak, 2017), we focus on $\phi(x) = \tanh(x)$, but we show that qualitatively similar dynamical regimes appear in network models with more realistic, positively defined activation functions (Figure S7). The transformed variable $\phi(x_i(t))$ is interpreted as the firing rate of unit *i*, and is also referred to as the *activity* variable.

The time evolution is specified by the following dynamics:

$$\dot{x_i}(t) = -x_i(t) + \sum_{j=1}^{N} J_{ij}\phi(x_j(t)) + I_i. \tag{6}$$

We considered a particular class of connectivity matrices, which can be written as a sum of two terms:

$$J_{ij} = g\chi_{ij} + P_{ij}. \tag{7}$$

Similarly to (Sompolinsky et al., 1988), χ_{ij} is a Gaussian all-to-all random matrix, where every element is drawn from a centered normal distribution with variance $1/N$. The parameter g scales the strength of random connections in the network, and we refer to it also as the *random strength*. The second term P_{ij} is a low-rank matrix. In this study, we consider the low-rank part of the connectivity fixed, while the random part varies between different realizations of the connectivity. Our results rely on two simplifying assumptions. The first one is that the low-rank term and the random term are statistically uncorrelated. The second one is that, as stated in Equation 8, the structured connectivity is weak in the large N limit, i.e., it scales as $1/N$, while the random connectivity components χ_{ij} scale as $1/\sqrt{N}$.

We first consider the simplest case where P_{ij} is a rank-one matrix, which can generally be written as the external product between two one-dimensional vectors m and n:

$$P_{ij} = \frac{m_i n_j}{N}. \tag{8}$$

According to our first assumption, the entries of vectors m and n are independent of the random bulk of the connectivity χ_{ij}. Note that the only non-zero eigenvalue of P is given by the scalar product $m^T n/N$, and the corresponding right and left eigenvectors are, respectively, vectors m and n. In the following, we will refer to the eigenvalue $m^T n/N$ as the *strength of the connectivity structure*, and to m and n as the *right- and left-connectivity vectors*. Here we focus on vectors obtained by generating the components from a joint Gaussian distribution.

More general connectivity structures of rank $r \ll N$ can be written as a sum of unit-rank terms

$$P_{ij} = \frac{m_i^{(1)} n_j^{(1)}}{N} + \ldots + \frac{m_i^{(r)} n_j^{(r)}}{N}, \tag{9}$$

and are therefore specified by r pairs of vectors $m^{(k)}$ and $n^{(k)}$, where different m vectors are linearly independent, and similarly for n vectors.

Overview of Dynamical Mean-Field Theory

Our results rely on a mathematical analysis of network dynamics based on Dynamical Mean-Field (DMF) theory (Sompolinsky et al., 1988; Rajan et al., 2010; Kadmon and Sompolinsky, 2015). To help navigate the analysis, here we provide first a succint overview of the approach. Full details are given further down in the section *Details of Dynamical Mean-Field Theory*.

DMF theory allows one to derive an effective description of the dynamics by averaging over the disorder originating from the random part of the connectivity. Across different realizations of the random connectivity matrix χ_{ij}, the sum of inputs to unit i is approximated by a Gaussian stochastic process $\eta_i(t)$

$$\sum_{j=1}^{N} J_{ij}\phi(x_j(t)) + I_i \approx \eta_i(t), \tag{10}$$

so that each unit obeys a Langevin-like equation:

$$\dot{x_i}(t) = -x_i(t) + \eta_i(t). \tag{11}$$

The Gaussian processes η_i can in principle have different first and second-order statistics for each unit, but are otherwise statistically independent across different units. As a consequence, the activations x_i of different units are also independent Gaussian stochastic processes, coupled only through their first and second-order statistics. The core of DMF theory consists of self-consistent equations for the mean μ_i and auto-correlation function $\Delta_i^I(t)$.

At equilibrium (i.e., in absence of transient dynamics) the equation for the mean μ_i of x_i is obtained by directly averaging Equation 6 over the random part of the connectivity. For a unit-rank connectivity, it reads

$$\mu_i = \kappa m_i + I_i, \tag{12}$$

where

$$\kappa = \frac{1}{N}\sum_{j=1}^{N} n_j[\phi_j]. \tag{13}$$

In the last equation, we adopted the short-hand notation $\phi_i(t) := \phi(x_i(t))$. Here $[\phi_j]$ is the average firing rate of unit j, i.e., $\phi(x_j)$ averaged over the Gaussian variable x_j. In a geometrical interpretation, the quantity κ represents the overlap between the left-connectivity vector n and the vector of average firing rates. Equivalently, it is given by a population average of $n_j[\phi_j]$, which can also be expressed as

$$\kappa = \int \mathrm{d}m\, \mathrm{d}n\, \mathrm{d}I\, p(m,n,I) n \int \mathcal{D}z \phi\left(m\kappa + I + \sqrt{\Delta_0^I} z\right) \tag{14}$$

where $p(m,n,I)$ is the joint distribution of components of vectors m, n and I. Δ_0^I is the variance of x_i (see below), and $\int \mathcal{D}z = \int_{-\infty}^{+\infty} e^{-z^2/2}/\sqrt{2\pi}\mathrm{d}z$.

The auto-correlation function $\Delta_i^I(t)$ quantifies the fluctuations of the activation x_i around the expected mean. Computing this auto-correlation function shows that it is identical for all units in the network, i.e., independent of i (see Equation 27). It can be decomposed into a static variance, which quantifies the fluctuations of the equilibrium values of x_i across different realizations of the random component of the connectivity, and an additional temporal variance which is present when the network is in a temporally fluctuating, chaotic state. In a stationary state, the variance $\Delta_0^I \equiv \Delta^I(t=0)$ can be expressed as

$$\Delta_0^I = g^2 \frac{1}{N}\sum_{j=1}^{N} [\phi_i^2]. \tag{15}$$

where $[\phi_i^2]$ is the average of $\phi_i^2(x)$ over the Gaussian variable x_i.

The right-hand-sides of Equations 13 and 15 show that both the mean μ_i and variance Δ_0^I depend on population-averaged, macroscopic quantities. To fully close the DMF description, the equations for single-unit statistics need to be averaged over the population. For static equilibrium dynamics, this leads to two coupled equations for two macroscopic quantities, the overlap κ and the static, population-averaged variance Δ_0:

$$\begin{aligned} \kappa &= F(\kappa, \Delta_0) \\ \Delta_0 &= G(\kappa, \Delta_0). \end{aligned} \tag{16}$$

Here F and G are two non-linear functions, the specific form of which depends on the geometrical arrangement of the connectivity vectors m and n and the input vector I. For temporally fluctuating, chaotic dynamics an additional macroscopic quantity (corresponding to the temporal variance) needs to be taken into account. In that case, the full DMF description is given by a system of three non-linear equations for three unknowns. The equilibrium states of the network dynamics are therefore obtained by solving these systems of equations using standard non-linear methods.

To describe the transient dynamics and assess the stability of the obtained equilibrium states, we determined the spectrum of eigenvalues at the obtained equilibrium fixed points. This spectrum consists of two components: a continuous, random component distributed within a circle in the complex plane, and a single outlier induced by the structured part of the connectivity (Figures S1A and S1D). The radius of the continuous component and the value of the outlier depend on the connectivity parameters. Although the two quantities in general are non-trivially coupled, the value of the radius is mostly controlled by the strength of the disorder, while the value of the outlier increases with the strength $m^T n/N$ of the rank-one structure (Figure S1F). The equilibrium is stable as long as the real part of all eigenvalues is less than unity. For large connectivity structure strengths, the outlier crosses unity, generating an instability that leads to the appearance of one-dimensional structured activity. Increasing the disorder strength on the other hand leads to another instability, corresponding to the radius of the continuous component crossing unity. This instability gives rise to chaotic, fluctuating activity.

When a linear readout with weights w_i is added to the network, its average output is given by

$$z(t) = \frac{1}{N}\sum_{i=1}^{N} w_i[\phi_i(t)], \tag{17}$$

i.e., by the projection of the average network firing rate on the readout vector w. This quantity is analogous to κ, except that the vector n is replaced by the vector w, so that similarly to Equation 14, the average readout can also be expressed as

$$z = \int \mathrm{d}m\, \mathrm{d}w\, \mathrm{d}I\, p(m,w,I)\, w \int \mathcal{D}y \phi\left(mz + I + \sqrt{\Delta_0^I}\, y\right) \tag{18}$$

and therefore directly depends on the joint distribution $p(m,w,I)$ which characterizes the geometric arrangement of vectors m, w and I.

The DMF theory can be directly extended to connectivity structures of rank r greater than one. The equilibrium mean input to unit i is then given by

$$\mu_i = \sum_{k=1}^{r} \kappa^{(k)} m_i^{(k)} + I_i. \tag{19}$$

The activity therefore lives in an $(r+1)$-dimensional space determined by the r right-connectivity vectors $m^{(k)}$ and the input vector I. It is characterized by r overlaps $\kappa^{(k)}$, each of which quantifies the amount of activity along the corresponding direction $m^{(k)}$. Averaging over the population, the DMF theory then leads to a system of $r+1$ nonlinear coupled equations for describing stationary dynamics.

DETAILS OF DYNAMICAL MEAN-FIELD THEORY

Here we provide the full details of the mathematical analysis. We start by examining the activity of a network with a rank-one structure in absence of external inputs ($I_i = 0\ \forall i$ in Equation 6).

Single-unit equations for spontaneous dynamics

We start by determining the statistics of the effective noise η_i to unit i, defined by

$$\eta_i(t) = g\sum_{j=1}^{N} \chi_{ij}\phi(x_j(t)) + \frac{m_i}{N}\sum_{j=1}^{N} n_j \phi(x_j(t)). \tag{20}$$

The DMF theory relies on the hypothesis that a disordered component in the coupling structure, here represented by χ_{ij}, efficiently decorrelates single neuron activity when the network is sufficiently large. We will show that this hypothesis of decorrelated activity is self-consistent for the specific network architecture we study.

As in standard DMF derivations, we characterize self-consistently the distribution of η_i by averaging over different realizations of the random matrix χ_{ij} (Sompolinsky et al., 1988; Rajan et al., 2010). In the following, $[.]$ indicates an average over the realizations of the random matrix χ_{ij}, while $\langle . \rangle$ stands for an average over different units of the network. Note that the network activity can be equivalently characterized in terms of input current variables $x_i(t)$ or their non-linear transforms $\phi(x_i(t))$. As these two quantities are not independent, the statistics of the distribution of the latter can be written in terms of the statistics of the former.

The mean of the effective noise received by unit i is given by:

$$[\eta_i(t)] = g\sum_{j=1}^{N} [\chi_{ij}\phi(x_j(t))] + \frac{m_i}{N}\sum_{j=1}^{N} n_j [\phi(x_j(t))]. \tag{21}$$

Under the hypothesis that in large networks, neural activity decorrelates (more specifically, that activity $\phi(x_j(t))$ is independent of its outgoing weights), we have:

$$[\eta_i(t)] = g\sum_{j=1}^{N}[\chi_{ij}][\phi(x_j(t))] + \frac{m_i}{N}\sum_{j=1}^{N} n_j[\phi(x_j(t))] = m_i\kappa \tag{22}$$

as $[\chi_{ij}] = 0$. Here we introduced

$$\kappa := \frac{1}{N}\sum_{j=1}^{N} n_j[\phi(x_j(t))] = \langle n_j[\phi_j(t)]\rangle, \tag{23}$$

which quantifies the overlap between the mean population activity vector and the left-connectivity vector n.

Similarly, the noise correlation function is given by

$$[\eta_i(t)\eta_j(t+\tau)] = g^2\sum_{k=1}^{N}\sum_{l=1}^{N}[\chi_{ik}\chi_{jl}][\phi(x_k(t))\phi(x_l(t+\tau))] + \frac{m_i m_j}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N} n_k n_l[\phi(x_k(t))\phi(x_l(t+\tau))]. \tag{24}$$

Note that every cross-term in the product vanishes since $[\chi_{ij}] = 0$. Similarly to standard DMF derivations (Sompolinsky et al., 1988), the first term on the r.h.s. vanishes for cross-correlations $(i \neq j)$ while it survives in the auto-correlation function $(i = j)$, as $[\chi_{ik}\chi_{jl}] = \delta_{ij}\delta_{kl}/N$. We get:

$$[\eta_i(t)\eta_j(t+\tau)] = \delta_{ij}g^2\langle[\phi_i(t)\phi_i(t+\tau)]\rangle + \frac{m_i m_j}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N} n_k n_l[\phi(x_k(t))\phi(x_l(t+\tau))]. \tag{25}$$

We focus now on the second term in the right-hand side. The corresponding sum contains N terms where $k = l$. This contribution vanishes in the large N limit because of the $1/N^2$ scaling. According to our starting hypothesis, when $k \neq l$, activity decorrelates: $[\phi_k(t)\phi_l(t+\tau)] = [\phi_k(t)][\phi_l(t+\tau)]$. To the leading order in N, we get:

$$\begin{aligned}[\eta_i(t)\eta_j(t+\tau)] &= \delta_{ij}g^2\langle[\phi_i(t)\phi_i(t+\tau)]\rangle + \frac{m_i m_j}{N^2}\sum_{k} n_k[\phi(x_k(t))]\sum_{l\neq k} n_l[\phi(x_l(t+\tau))] \\ &= \delta_{ij}g^2\langle[\phi_i(t)\phi_i(t+\tau)]\rangle + m_i m_j\kappa^2\end{aligned} \tag{26}$$

so that:

$$[\eta_i(t)\eta_j(t+\tau)] - [\eta_i(t)][\eta_j(t)] = \delta_{ij}g^2\langle[\phi_i(t)\phi_i(t+\tau)]\rangle. \tag{27}$$

We therefore find that the statistics of the effective input are uncorrelated across different units, so that our initial hypothesis is self-consistent.

To conclude, for every unit i, we computed the first- and the second-order statistics of the effective input $\eta_i(t)$. The expressions we obtained show that the individual noise statistics depend on the statistics of the full network activity. In particular, the mean of the effective input depends on the average overlap κ, but varies from unit to unit through the components of the right-connectivity vector m. On the other hand, the auto-correlation of the effective input is identical for all units, and determined by the population-averaged firing rate auto-correlation $\langle[\phi_i(t)\phi_i(t+\tau)]\rangle$.

Once the statistics of $\eta_i(t)$ have been determined, a self-consistent solution for the activation variable $x_i(t)$ can be derived by solving the Langevin-like stochastic process from Equation 11. As a first step, we look at its stationary solutions, which correspond to the fixed points of the original network dynamics.

Population-averaged equations for stationary solutions

For any solution that does not depend on time, the mean μ_i and the variance Δ_0^I of the variable x_i with respect to different realizations of the random connectivity coincide with the statistics of the effective noise η_i. From Equations 22 and 27, the mean μ_i and variance Δ_0^I of the input to unit i therefore read

$$\begin{aligned}\mu_i &:= [x_i] = m_i\kappa \\ \Delta_0^I &:= [x_i^2] - [x_i]^2 = g^2\langle[\phi_i^2]\rangle\end{aligned} \tag{28}$$

while any other cross-variance $[x_i x_j] - [x_i][x_j]$ vanishes. We conclude that, on average, the structured connectivity P_{ij} shapes the network activity along the direction specified by its right eigenvector m. Such a heterogeneous stationary state critically relies on a non-vanishing overlap κ between the left eigenvector n and the average population activity vector $[\phi]$. Across different realizations of the random connectivity, the input currents x_i fluctuate around these mean values. The typical size of fluctuations is determined by the individual variance Δ_0^I, equal for every unit in the network.

The r.h.s. of Equation 28 contains two population averaged quantities, the overlap κ and the second moment of the activity $\langle[\phi_i^2]\rangle$. To close the equations, these quantities need to be expressed self-consistently. Averaging Equation 28 over the population, we get expressions for the population-averaged mean μ and variance Δ_0 of the input:

$$\begin{aligned}\mu &:= \langle[x_i]\rangle = \langle m_i\rangle\kappa \\ \Delta_0 &:= \left\langle\left[x_i^2\right]\right\rangle - \langle[x_i]\rangle^2 = g^2\left\langle\left[\phi_i^2\right]\right\rangle + \left(\left\langle m_i^2\right\rangle - \langle m_i\rangle^2\right)\kappa^2.\end{aligned} \tag{29}$$

Note that the total population variance Δ_0 is a sum of two terms: the first term, proportional to the strength of the random part of connectivity, coincides with the individual variability Δ_0^I which emerges from different realizations of χ_{ij}; the second term, proportional to the variance of the right-connectivity vector m, coincides with the variance induced at the population level by the spread of the mean values $\mu_i \propto m_i$. When the vector m is homogeneous ($m_i = \overline{m}$), input currents x_i are centered around the same mean value μ, and the second variance term vanishes.

We next derive appropriate expression for the r.h.s. terms κ and $\langle[\phi_i^2]\rangle$. To start with, we rewrite $[\phi_i]$ by substituting the average over the random connectivity with the equivalent Gaussian integral:

$$[\phi_i] = \int \mathcal{D}z\phi\left(\mu_i + \sqrt{\Delta_0^I}z\right) \tag{30}$$

where we used the short-hand notation $\int \mathcal{D}z = \int_{-\infty}^{+\infty} e^{-z^2/2}/\sqrt{2\pi}\mathrm{d}z$. To obtain κ, $[\phi_i]$ needs to be multiplied by n_i and averaged over the population. This average can be expressed by representing the fixed vectors m and n through the joint distribution of their elements over the components:

$$p(m,n) = \frac{1}{N}\sum_{j=1}^{N}\delta(m - m_j)\delta(n - n_j). \tag{31}$$

This leads to

$$\begin{aligned}\kappa &= \left\langle n_i \int \mathcal{D}z\phi\left(\mu_i + \sqrt{\Delta_0^I}z\right)\right\rangle \\ &= \int \mathrm{d}m \int \mathrm{d}n\, p(m,n)\, n \int \mathcal{D}z\phi\left(m\kappa + \sqrt{\Delta_0^I}z\right).\end{aligned} \tag{32}$$

Similarly, a suitable expression for the second-order momentum of the firing rate is given by:

$$\left\langle\left[\phi_i^2\right]\right\rangle = \int \mathrm{d}m\, p(m) \int \mathcal{D}z\phi^2\left(m\kappa + \sqrt{\Delta_0^I}z\right). \tag{33}$$

Equations 32 and 33, combined with Equation 29, provide a closed set of equations for determining κ and Δ_0 once the vectors m and n have been specified.

To further simplify the problem, we reduce the full distribution $p(m,n)$ of elements m_i and n_i to their first- and second-order momenta. That is equivalent to substituting the probability density $p(m,n)$ with a bivariate Gaussian distribution. We therefore write:

$$\begin{aligned}m &= M_m + \Sigma_m\sqrt{1-\rho}\,x_1 + \Sigma_m\sqrt{\rho}\,y \\ n &= M_n + \Sigma_n\sqrt{1-\rho}\,x_2 + \Sigma_n\sqrt{\rho}\,y\end{aligned} \tag{34}$$

where x_1, x_2 and y are three normal Gaussian processes. Here, M_m (resp. M_n) and Σ_m (resp. Σ_n) correspond to the mean and the standard deviation of m (resp. n), while the covariance between m and n is given by $\langle m_i n_i\rangle - M_m M_n = \Sigma_m\Sigma_n\rho$. Within a geometrical interpretation, M_m and M_n are the projections of N–dimensional vectors m and n onto the unitary vector $u = (1,1,\ldots 1)/N$, $\Sigma_m\sqrt{\rho}$ and $\Sigma_n\sqrt{\rho}$ are the projections onto a direction orthogonal to u and common to m and n, and $\Sigma_m\sqrt{1-\rho}$ and $\Sigma_n\sqrt{1-\rho}$ scale the parts of m and n that are mutually orthogonal.

The expression for κ becomes:

$$\kappa = \int \mathcal{D}y \int \mathcal{D}x_2 \left(M_n + \Sigma_n\sqrt{1-\rho}\,x_2 + \Sigma_n\sqrt{\rho}\,y\right) \int \mathcal{D}z \int \mathcal{D}x_1\, \phi\left(\kappa\left(M_m + \Sigma_m\sqrt{1-\rho}\,x_1 + \Sigma_m\sqrt{\rho}\,y\right) + \sqrt{\Delta_0^I}z\right) \tag{35}$$

which gives rise to three terms when expanding the sum $M_n + \Sigma_n\sqrt{1-\rho}x_2 + \Sigma_n\sqrt{\rho}y$. The first term can be rewritten as:

$$\begin{aligned}&M_n \int \mathcal{D}z\,\phi\left(M_m\kappa + \sqrt{\Delta_0^I + \Sigma_m^2\kappa^2}\,z\right) \\ &= M_n \int \mathcal{D}z\,\phi\left(\mu + \sqrt{\Delta_0}z\right) \\ &= M_n\langle[\phi_i]\rangle,\end{aligned} \tag{36}$$

which coincides with the overlap between vectors n and $[\phi]$ along the unitary direction $u = (1, 1, \ldots 1)/N$. In the last step, we rewrote our expression for κ in terms of the population averaged statistics μ and Δ_0 (Equation 29).

The second term vanishes, while the third one gives:

$$\Sigma_n \sqrt{\rho} \int \mathcal{D}y\, y \int \mathcal{D}z \int \mathcal{D}x_1\, \phi\left(\kappa\left(M_m + \Sigma_m \sqrt{1-\rho}\, x_1 + \Sigma_m \sqrt{\rho}\, y\right) + \sqrt{\Delta_0^l}\, z\right) = \kappa \rho \Sigma_m \Sigma_n \left\langle \left[\phi_i^{'}\right] \right\rangle \tag{37}$$

which coincides with the overlap between n and $[\phi]$ in a direction orthogonal to u. Here we used the equality:

$$\int \mathcal{D}z\, z f(z) = \int \mathcal{D}z\, \frac{df(z)}{dz} \tag{38}$$

which is obtained by integrating by parts.

Through a similar reasoning we obtain:

$$\left\langle \left[\phi_i^2\right] \right\rangle = \int \mathcal{D}z\, \phi^2\left(\mu + \sqrt{\Delta_0}\, z\right) \tag{39}$$

as in standard DMF derivations.

To conclude, the mean-field description of stationary solutions reduces to the system of three implicit equations for μ, κ and Δ_0:

$$\begin{aligned} \mu &= M_m \kappa \\ \Delta_0 &= g^2 \left\langle \left[\phi_i^2\right] \right\rangle + \Sigma_m^2 \kappa^2 \\ \kappa &= M_m \left\langle [\phi_i] \right\rangle + \kappa \rho \Sigma_m \Sigma_n \left\langle \left[\phi_i^{'}\right] \right\rangle. \end{aligned} \tag{40}$$

Both averages $\langle [.] \rangle$ are performed with respect to a Gaussian distribution of mean μ and variance Δ_0. Once μ, Δ_0 and κ have been determined, the single unit mean μ_i and the individual variance Δ_0^l are obtained from Equation 28.

The dynamical mean-field equations given in Equation 40 can be fully solved to determine stationary solutions. Detailed descriptions of these solutions are provided further down for two particular cases: (i) overlap between m and n only along the unitary direction u ($M_m \neq 0$, $M_n \neq 0$, $\rho = 0$); (ii) overlap between m and n only in a direction orthogonal to u ($M_m = M_n = 0$, $\rho \neq 0$).

Transient dynamics and stability of stationary solutions

We now turn to transient dynamics around fixed points, and to the related problem of evaluating whether the stationary solutions found within DMF are stable with respect to the original network dynamics (Equation 6).

For any given realization of the connectivity matrix, the network we consider is completely deterministic. We can then study the local, transient dynamics by linearizing the dynamics around any stationary solution. We therefore look at the time evolution of a small displacement away from the fixed point: $x(t) = x_i^0 + x_i^1(t)$. For any generic stationary solution $\{x_i^0\}$ the linearized dynamics are given by the stability matrix S_{ij} which reads:

$$S_{ij} = \phi^{'}\left(x_j^0\right)\left(g\chi_{ij} + \frac{m_i n_j}{N}\right). \tag{41}$$

If the real part of every eigenvalue of S_{ij} is smaller than unity, the perturbation decays in time and thus the stationary solution is stable.

Homogeneous stationary solutions

We first consider homogeneous stationary solutions, for which $x_i^0 = \bar{x}$ for all units. A particular homogeneous solution is the trivial solution $\bar{x} = 0$, which the network admits for all parameter values when the transfer function is $\phi(x) = \tanh(x)$. Other homogeneous solutions can be obtained when the vector m is homogeneous, i.e., $m_i = M_m$ for all i.

For homogeneous solutions, the stability matrix reduces to a scaled version of the connectivity matrix:

$$S_{ij} = \phi'(\bar{x}) J_{ij}. \tag{42}$$

We are thus left with the problem of evaluating the eigenspectrum of the global connectivity matrix J_{ij}. The matrix J_{ij} consists of a full-rank component χ_{ij}, the entries of which are drawn at random, and of a structured component of small dimensionality with fixed entries. We focus on the limit of large networks; in that limit, an analytical prediction for the spectrum of its eigenvalues can be derived.

Because of the $1/N$ scaling, the matrix norm of P_{ij} is bounded as N increases. We can then apply results from random matrix theory (Tao, 2013) which predict that, in the large N limit, the eigenspectra of the random and the structured parts do not interact, but sum together. The eigenspectrum of J_{ij} therefore consists of two separated components, inherited respectively from the random and the structured terms (Figure S1A). Similarly to (Girko, 1985), the random term χ_{ij} returns a set of $N - 1$ eigenvalues which lie on the complex plane in a compact circular region of radius g. In addition to this component, the eigenspectrum of J_{ij} contains the non-zero

eigenvalues of P_{ij}: in the case of a rank-one matrix, one single outlier eigenvalue is centered at the position $\sum_i m_i n_i/N = \langle m_i n_i \rangle$. In Figure S1B we measure both the outlier position and the radius of the compact circular component. We show that deviations from the theoretical predictions are in general small and decay to zero as the system size is increased.

Going back to the stability matrix $S_{ij} = \phi'(\bar{x})J_{ij}$, we conclude that a homogeneous stationary solution can lose stability in two different ways, when either $m^T n/N$ or g become larger than $1/\phi'(\bar{x})$. We expect different kinds of instabilities to occur in the two cases. When g crosses the instability line, a large number of random directions become unstable at the same time. As in (Sompolinsky et al., 1988), this instability is expected to lead to the onset of irregular temporal activity. When the instability is lead by the outlier, instead, the trivial fixed point becomes unstable in one unique direction given by the corresponding eigenvector. When $g = 0$, this eigenvector coincides exactly with m. For finite values of the disorder g, the outlier eigenvector fluctuates depending on the random part of the connectivity, but remains strongly correlated with m (Figure S1C), which therefore determines the average direction of the instability. Above the instability, as the network dynamics is completely symmetric with respect to a change of sign of the input variables, we expect the non-linear boundaries to generate two symmetric stationary solutions.

Heterogeneous stationary solutions

A second type of possible stationary solutions are heterogeneous fixed points, in which different units reach different equilibrium values. For such fixed points, the linearized stability matrix S_{ij} is obtained by multiplying each column of the connectivity matrix J_{ij} by a different gain value (see Equation 41), so that the eigenspectrum of S_{ij} is not trivially related to the spectrum of J_{ij}.

Numerical investigations reveal that, as for J_{ij}, the eigenspectrum of S_{ij} consists of two discrete components: one compact set of $N-1$ eigenvalues contained in a circle on the complex plane, and a single isolated outlier eigenvalue (Figure S1D).

As previously noticed in (Harish and Hansel, 2015), the radius of the circular compact set r can be computed as in (Rajan and Abbott, 2006; Aljadeff et al., 2015b) by summing the variances of the distributions in every column of S_{ij}. To the leading order in N:

$$r = g\sqrt{\frac{1}{N}\sum_{j=1}^{N}\phi'^2\left(x_j^0\right)} \tag{43}$$

which, in large networks, can be approximated by the mean-field average:

$$r = g\sqrt{\langle[\phi_i'^2]\rangle}. \tag{44}$$

Note that, because of the weak scaling in P_{ij}, the structured connectivity term does not appear explicitly in the expression for the radius. As the structured part of the connectivity determines the heterogeneous fixed point, the value of r however depends implicitly on the structured connectivity term through $\langle[\phi_i'^2]\rangle$, which is computed as a Gaussian integral over a distribution with mean μ and variance Δ_0 given by Equation 40. In Figures S1D–S1F, we show that Equation 44 approximates well the radius of finite-size, numerically computed eigenspectra. Whenever the mean-field theory predicts instabilities led by r, we expect the network dynamics to converge to irregular non-stationary solutions. Consistently, at the critical point, where $r = 1$, the DMF equations predict the onset of temporally fluctuating solutions (see later on in STAR Methods).

We now turn to the problem of evaluating the position of the outlier eigenvalue. In the case of heterogeneous fixed points, the structured and the random components of the matrix S_{ij} are strongly correlated, as they both scale with the multiplicative factor $\phi'(x_j^0)$, which correlates with the particular realization of the random part of the connectivity χ_{ij}. As a consequence, χ_{ij} cannot be considered as a truly random matrix with respect to $m_i\phi'(x_j^0)n_j/N$, and in contrast to the case of homogeneous fixed points, results from (Girko, 1985) do not hold.

We determined numerically the position of the outlier in finite-size eigenspectra (Figures S1D–S1F). We found that its value indeed significantly deviates from the only non-zero eigenvalue of the rank-one structure $m_i\phi'(x_j^0)n_j/N$, which can be computed in the mean-field framework (when $\rho = 0$, it corresponds to $M_m M_n\langle[\phi_i']\rangle + M_n\kappa\Sigma_m^2\langle[\phi_i'']\rangle$). On the other hand, the value of the outlier coincides exactly with the eigenvalue of $m_i\phi'(x_j^0)n_j/N$ whenever the random component χ_{ij} is shuffled (black dots in Figure S1F). This observation confirms that the position of the outlier critically depends on the correlations existing between the rank-one structure $m_i\phi'(x_j^0)n_j/N$ and the specific realization of the random bulk χ_{ij}.

Mean-field analysis of transient dynamics and stability of stationary solutions

As for heterogeneous fixed points we were not able to assess the position of the outlying eigenvalue using random matrix theory, we turned to a mean-field analysis to determine transient activity. This analysis allowed us to determine accurately the position of the outlier, and therefore the stability of heterogeneous fixed points. The approach exploited here is based on (Kadmon and Sompolinsky, 2015).

We consider the stability of the single unit activation x_i when averaged across different realizations of the random connectivity and its random eigenmodes. Directly averaging across realizations the network dynamics defined in Equation 6 yields the time evolution of the mean activation μ_i of unit i:

$$\dot{\mu}_i(t) = -\mu_i(t) + m_i\kappa(t). \tag{45}$$

We observe that we can write: $\mu_i(t) = m_i\tilde{\kappa}(t)$, where $\tilde{\kappa}$ is the low-pass filtered version of κ: $(1 + \mathrm{d}/\mathrm{d}t)\tilde{\kappa}(t) = \kappa(t)$. Small perturbations around the fixed point solution read: $\mu_i(t) = \mu_i^0 + \mu_i^1(t)$. The equilibrium values μ_i^0 correspond to the DMF stationary solution computed from Equation 28 and 40: $\mu_i^0 = m_i\kappa^0$. The first-order perturbations thus obey:

$$\dot{\mu}_i^1(t) = -\mu_i^1(t) + m_i\kappa^1(t), \tag{46}$$

indicating that the decay timescale of the mean activity is inherited by the decay time constant of κ^1. An additional equation for the time evolution of κ^1 thus needs to be derived.

When activity is perturbed, the firing activity ϕ_i of unit i can be evaluated at the first order: $\phi_i^0 \to \phi_i^0 + \phi_i^1(t) = \phi(x_i^0) + \phi'(x_i^0)x_i^1(t)$. As a consequence, the first-order in κ reads:

$$\kappa^1(t) = \left\langle n_i\left[\phi'(x_i^0)x_i^1(t)\right]\right\rangle. \tag{47}$$

Summing Equation 47 to its time-derivative, we get:

$$\dot{\kappa}^1(t) = -\kappa^1(t) + \left(1 + \frac{\mathrm{d}}{\mathrm{d}t}\right)\left\langle n_i\left[\phi'(x_i^0)x_i^1(t)\right]\right\rangle. \tag{48}$$

In order to simplify the r.h.s., we start by considering the average with respect to the random part of the connectivity for a single unit i. In order to compute $[\phi'(x_i^0)x_i^1]$, we explicitly build x_i^0 and $x_i^t := x_i(t)$ as Gaussian variables centered respectively in μ_i^0 and μ_i^t. We will call Δ_0^{l0} and Δ_0^{lt} the variances of the two variables, and $\Delta^{l,t:0}$ their two-times correlation defined by $\Delta^{l,t:0} = [x_i^t x_i^0] - [x_i^t][x_i^0]$. We can then write the two variables as

$$\begin{aligned} x_i^0 &= \mu_i^0 + \sqrt{\Delta_0^{l0} - \Delta^{l,t:0}}x_1 + \sqrt{\Delta^{l,t:0}}y \\ x_i^t &= \mu_i^t + \sqrt{\Delta_0^{lt} - \Delta^{l,t:0}}x_2 + \sqrt{\Delta^{l,t:0}}y \end{aligned} \tag{49}$$

The first-order response of x_i is given by the difference between x_i^t and x_i^0, and reads:

$$x_i^1 = \mu_i^1 + \sqrt{\Delta_0^{lt} - \Delta^{l,t:0}}x_2 - \sqrt{\Delta_0^{l0} - \Delta^{l,t:0}}x_1. \tag{50}$$

As in classical DMF derivations (Sompolinsky et al., 1988; Rajan et al., 2010; Kadmon and Sompolinsky, 2015), x_1, x_2 and y are standard normal variables. By integrating over their distributions we can write:

$$\left[\phi'(x_i^0)x_i^1\right] = \int \mathcal{D}x_1 \int \mathcal{D}x_2 \left(\mu_i^1 + \sqrt{\Delta_0^{lt} - \Delta^{l,t:0}}x_2 - \sqrt{\Delta_0^{l0} - \Delta^{l,t:0}}x_1\right) \int \mathcal{D}y\phi'\left(\mu_i^0 + \sqrt{\Delta_0^{l0} - \Delta^{l,t:0}}x_1 + \sqrt{\Delta^{l,t:0}}y\right). \tag{51}$$

Integrating by parts as in Equation 38 we get:

$$\left[\phi'(x_i^0)x_i^1\right] = \mu_i^1\left[\phi_i'\right] + \left(\Delta^{l,t:0} - \Delta_0^{l0}\right)\left[\phi_i''\right] \tag{52}$$

where the Gaussian integrals $[\phi_i']$ and $[\phi_i'']$ are evaluated using the fixed point statistics.

Note that, at the fixed point, $\Delta^{l,t:0} = \Delta_0^{l0}$. As a consequence, $\Delta^{l,t:0} - \Delta_0^{l0}$ gives a first-order response:

$$\Delta^{l,1:0} := \Delta^{l,t:0} - \Delta_0^{l0} = \left[x_i^1 x_i^0\right] - \left[x_i^1\right]\left[x_i^0\right] = \left[x_i^1 x_i^0\right] - \mu_i^0\mu_i^1 \tag{53}$$

which can be rewritten as a function of the global second-order statistics $\Delta^{1:0} = \langle[x_i^1 x_i^0]\rangle - \langle[x_i^1]\rangle\langle[x_i^0]\rangle$ as:

$$\begin{aligned} \Delta^{l,1:0} &= \Delta^{1:0} - \left\{\left\langle\mu_i^1\mu_i^0\right\rangle - \left\langle\mu_i^1\right\rangle\left\langle\mu_i^0\right\rangle\right\} \\ &= \Delta^{1:0} - \Sigma_m^2\tilde{\kappa}^0\tilde{\kappa}^1. \end{aligned} \tag{54}$$

Equation 54 can be rewritten in terms of the first-order perturbation for the global equal-time variance: $\Delta_0^1 = \Delta_0^t - \Delta_0^0$. We consider that, by definition:

$$\begin{aligned} \Delta^{1:0} &= \sum_{j=1}^{N} x_j^1 \frac{\partial\Delta^{t:0}}{\partial x_j^t}\bigg|_0 \\ \Delta_0^1 &= \sum_{j=1}^{N} x_j^1 \frac{\partial\Delta_0^t}{\partial x_j^t}\bigg|_0. \end{aligned} \tag{55}$$

We then observe that, when the derivatives are evaluated at the fixed point, we have:

$$\left.\frac{\partial \Delta^{t:0}}{\partial x_j^t}\right|_0 = \frac{1}{2}\left.\frac{\partial \Delta_0^t}{\partial x_j^t}\right|_0, \tag{56}$$

and we conclude that:

$$\Delta^{1:0} = \frac{1}{2}\Delta_0^1 \tag{57}$$

Equation 52 thus becomes:

$$\left[\phi'(x_i^0)x_i^1\right] = m_i\,\tilde{\kappa}^1\left[\phi_i'\right] + \left(\frac{\Delta_0^1}{2} - \Sigma_m^2\tilde{\kappa}^0\tilde{\kappa}^1\right)\left[\phi_i''\right]. \tag{58}$$

In a second step, we perform the average across different units of the population, by writing m and n as in Equation 34. After some algebra, we get:

$$\begin{aligned}\left\langle n_i\left[\phi'(x_i^0)x_i^1(t)\right]\right\rangle &= \tilde{\kappa}^1\left[(M_mM_n + \rho\Sigma_m\Sigma_n)\left\langle\left[\phi_i'\right]\right\rangle + \rho\kappa^0 M_m\Sigma_m\Sigma_n\left\langle\left[\phi_i''\right]\right\rangle\right] + \frac{\Delta_0^1}{2}\left[M_n\left\langle\left[\phi_i''\right]\right\rangle + \rho\kappa^0\Sigma_m\Sigma_n\left\langle\left[\phi_i'''\right]\right\rangle\right] \\ &:= \tilde{\kappa}^1 a + \Delta_0^1 b\end{aligned} \tag{59}$$

where constants a and b were defined as:

$$\begin{aligned}a &= (M_mM_n + \rho\Sigma_m\Sigma_n)\left\langle\left[\phi_i'\right]\right\rangle + \rho\kappa^0 M_m\Sigma_m\Sigma_n\left\langle\left[\phi_i''\right]\right\rangle \\ b &= \frac{1}{2}\left\{M_n\left\langle\left[\phi_i''\right]\right\rangle + \rho\kappa^0\Sigma_m\Sigma_n\left\langle\left[\phi_i'''\right]\right\rangle\right\}.\end{aligned} \tag{60}$$

The time evolution of κ can be finally rewritten as:

$$\dot{\kappa}^1(t) = -\kappa^1(t) + \left(1 + \frac{\mathrm{d}}{\mathrm{d}t}\right)\left\{\tilde{\kappa}^1 a + \Delta_0^1 b\right\}, \tag{61}$$

so that the time evolution of the perturbed variance must be considered as well.

In order to isolate the evolution law of Δ_0, we rewrite the activation variable $x_i(t)$ by separating the uniform and the heterogeneous components: $x_i(t) = \mu(t) + \delta x_i(t)$. The time evolution for the residual $\delta x_i(t)$ is given by:

$$\delta\dot{x}_i(t) = -\delta x_i(t) + g\sum_{j=1}^{N}\chi_{ij}\phi(x_j(t)) + (m_i - M_m)\kappa(t) \tag{62}$$

so that, squaring:

$$\left(\frac{\mathrm{d}\delta x_i(t)}{\mathrm{d}t}\right)^2 + 2\delta x_i(t)\frac{\mathrm{d}\delta x_i(t)}{\mathrm{d}t} + \delta x_i(t)^2 = g^2\sum_{j=1}^{N}\sum_{k=1}^{N}\chi_{ij}\chi_{ik}\phi(x_j(t))\phi(x_k(t)) + (m_i - M_m)^2\kappa(t)^2 + g(m_i - M_m)\kappa(t)\sum_{k=1}^{N}\chi_{ij}\phi(x_k(t)). \tag{63}$$

Averaging over i and the realizations of the disorder yields:

$$\begin{aligned}\frac{\mathrm{d}\Delta_0(t)}{\mathrm{d}t} &= -\Delta_0(t) + g^2\left\langle\left[\phi_i^2(t)\right]\right\rangle + \Sigma_m^2\kappa(t)^2 - \left\langle\left[\left(\frac{\mathrm{d}\delta x_i(t)}{\mathrm{d}t}\right)^2\right]\right\rangle \\ &:= -\Delta_0(t) + G(\mu, \Delta_0, \kappa) - \left\langle\left[\left(\frac{\mathrm{d}\delta x_i(t)}{\mathrm{d}t}\right)^2\right]\right\rangle\end{aligned} \tag{64}$$

as by definition we have: $\langle[\delta x_i^2(t)]\rangle = \Delta_0(t)$.

Expanding the dynamics of Δ_0 to the first order, we get:

$$\dot{\Delta}_0^1(t) = -\Delta_0^1(t) + \mu^1\left.\frac{\partial G}{\partial \mu}\right|_0 + \Delta_0^1\left.\frac{\partial G}{\partial \Delta_0}\right|_0 + \kappa^1\left.\frac{\partial G}{\partial \kappa}\right|_0. \tag{65}$$

Note that we could neglect the contributions originating from the last term of Equation 64 because they do not enter at the leading order. Indeed we have:

$$\frac{\partial}{\partial\mu}\left\langle\left[\left(\frac{\mathrm{d}\delta x_i(t)}{\mathrm{d}t}\right)^2\right]\right\rangle\Big|_0 = 2\left\langle\left[\frac{\mathrm{d}\delta x_i(t)}{\mathrm{d}t}\frac{\partial}{\partial\mu}\frac{\mathrm{d}\delta x_i(t)}{\mathrm{d}t}\right]\right\rangle\Big|_0 = 0 \tag{66}$$

since temporal derivatives for every i vanish when evaluated at the fixed point.

A little algebra returns the last three linear coefficients:

$$\begin{aligned}\frac{\partial G}{\partial\mu}\Big|_0 &= 2g^2\left\langle\left[\phi_i\phi_i'\right]\right\rangle\\ \frac{\partial G}{\partial\Delta_0}\Big|_0 &= g^2\left\{\left\langle\left[\phi_i'^2\right]\right\rangle + \left\langle\left[\phi_i\phi_i''\right]\right\rangle\right\}\\ \frac{\partial G}{\partial\kappa}\Big|_0 &= 2\Sigma_m^2\kappa^0.\end{aligned} \tag{67}$$

Collecting all the results together in Equation 61 we obtain:

$$\dot{\kappa}^1(t) = -\kappa^1(t) + a\kappa^1(t) + b\left\{\mu^1\frac{\partial G}{\partial\mu}\Big|_0 + \Delta_0^1\frac{\partial G}{\partial\Delta_0}\Big|_0 + \kappa^1\frac{\partial G}{\partial\kappa}\Big|_0\right\}. \tag{68}$$

By averaging Equation 45 we furthermore obtain:

$$\dot{\mu}^1(t) = -\mu^1(t) + M_m\kappa^1. \tag{69}$$

We finally obtained that the perturbation timescale is determined by the population-averaged dynamics:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\mu^1\\ \Delta_0^1\\ \kappa^1\end{pmatrix} = -\begin{pmatrix}\mu^1\\ \Delta_0^1\\ \kappa^1\end{pmatrix} + \mathcal{M}\begin{pmatrix}\mu^1\\ \Delta_0^1\\ \kappa^1\end{pmatrix} \tag{70}$$

where the evolution matrix $\mathcal{M}$ is defined as:

$$\mathcal{M} = \begin{pmatrix}0 & 0 & M_m\\ 2g^2\left\langle\left[\phi_i\phi_i'\right]\right\rangle & g^2\left\{\left\langle\left[\phi_i'^2\right]\right\rangle + \left\langle\left[\phi_i\phi_i''\right]\right\rangle\right\} & 2\Sigma_m^2\kappa^0\\ 2bg^2\left\langle\left[\phi_i\phi_i'\right]\right\rangle & bg^2\left\{\left\langle\left[\phi_i'^2\right]\right\rangle + \left\langle\left[\phi_i\phi_i''\right]\right\rangle\right\} & b2\Sigma_m^2\kappa^0 + a\end{pmatrix}. \tag{71}$$

Note that one eigenvalue of matrix $\mathcal{M}$, which corresponds to the low-pass filtering between κ and μ, is always fixed to zero.

Equations 70 and 71 reveal that, during the relaxation to equilibrium, the transient dynamics of the first- and second-order statistics of the activity are tightly coupled. Diagonalizing $\mathcal{M}$ allows to retrieve the largest decay timescale of the network, which indicates the average, structural stability of stationary states.

When an outlier eigenvalue is present in the eigenspectrum of the stability matrix S_{ij}, the largest decay timescale from $\mathcal{M}$ predicts its position. The corresponding eigenvector $\hat{e}$ contains indeed a structured component along m, which is not washed out by averaging across different realizations of χ_{ij}.

The second non-zero eigenvalue of $\mathcal{M}$, which vanishes at $g = 0$, measures a second and smaller effective timescale, which derives from averaging across the remaining $N-1$ random modes.

Varying g, we computed the largest eigenvalue of $\mathcal{M}$ for corresponding stationary solutions of mean-field equations. In Figure S1F we show that, when the stability eigenspectrum includes an outlier eigenvalue, its position is correctly predicted by the largest eigenvalue of $\mathcal{M}$. The mismatch between the two values is small and can be understood as a finite-size effect (Figure S1E, gray).

To conclude, we found that the stability of arbitrary stationary solutions can be assessed by evaluating, with the help of mean-field theory, both the values of the radius (Equation 44) and the outlier (Equation 71) of the stability eigenspectrum. Instabilities led by the two different components are expected to reshape activity into two qualitatively different classes of dynamical regimes, which are discussed in detail, further in STAR Methods, for two specific classes of structures.

Dynamical Mean Field equations for chaotic solutions

When a stationary state loses stability due to the compact component of the stability eigenspectrum, the network activity starts developing irregular temporal fluctuations. Such temporally fluctuating states can be described within the DMF theory by taking into account the full temporal auto-correlation function of the effective noise η_i (Sompolinsky et al., 1988). For the sake of simplicity, here we derive directly the mean-field equations for population-averaged statistics, and we eventually link them back to single unit quantities.

By differentiating twice Equation 11, and by substituting the appropriate expression for the statistics of the noise η_i, we derive that the auto-correlation function $\Delta(\tau) = \langle [x_i(t+\tau)x_i(t)] \rangle - \langle [x_i(t)] \rangle^2$ obeys the second-order differential equation:

$$\ddot{\Delta}(\tau) = \Delta(\tau) - g^2 \langle [\phi_i(t)\phi_i(t+\tau)] \rangle - \Sigma_m^2 \kappa^2. \tag{72}$$

In this context, the activation variance Δ_0 coincides with the peak of the full auto-correlation function: $\Delta_0 = \Delta(\tau = 0)$. We expect the total variance to include a temporal term, coinciding with the amplitude of chaotic fluctuations, and a quenched one, representing the spread across the population due to the disorder in χ_{ij} and the structure imposed by the right-connectivity vector m.

In order to compute the full rate auto-correlation function $\langle [\phi_i(t)\phi_i(t+\tau)] \rangle$, we need to explicitly build two correlated Gaussian variables $x(t)$ and $x(t+\tau)$, such that:

$$\begin{aligned} &\langle [x_i(t)] \rangle = \langle [x_i(t+\tau)] \rangle = \mu \\ &\left\langle \left[x_i^2(t)\right] \right\rangle - \langle [x_i(t)] \rangle^2 = \left\langle \left[x_i^2(t+\tau)\right] \right\rangle - \langle [x_i(t)] \rangle^2 = \Delta_0 \\ &\langle [x_i(t+\tau)x_i(t)] \rangle - \langle [x_i(t)] \rangle^2 = \Delta(\tau). \end{aligned} \tag{73}$$

Following previous studies (Sompolinsky et al., 1988; Rajan et al., 2010), we obtain:

$$\langle [\phi_i(t)\phi_i(t+\tau)] \rangle = \int \mathcal{D}z \left[\int \mathcal{D}x \phi\left(\mu + \sqrt{\Delta_0 - \Delta}\, x + \sqrt{\Delta}\, z\right) \right]^2 \tag{74}$$

where we used the short-hand notation $\Delta := \Delta(\tau)$ and we assumed for simplicity $\Delta > 0$. As we show later, this requirement is satisfied by our final solution.

In order to visualize the dynamics of the solutions of Equation 72, we study the equivalent problem of a classical particle moving in a one-dimensional potential (Sompolinsky et al., 1988; Rajan et al., 2010):

$$\ddot{\Delta}(\tau) = -\frac{\partial V}{\partial \Delta} \tag{75}$$

where the potential V is given by an integration over Δ:

$$V(\Delta, \Delta_0) = -\frac{\Delta^2}{2} + g^2 \langle [\Phi_i(t)\Phi_i(t+\tau)] \rangle + \Sigma_m^2 \kappa^2 \Delta \tag{76}$$

and $\Phi(x) = \int_{-\infty}^{x} \phi(x')\mathrm{d}x'$. As the potential V depends self-consistently on the initial condition Δ_0, the shape of the auto-correlation function $\Delta(\tau)$ depends parametrically on the value of Δ_0. Similarly to previous works, we isolate the solutions that decay monotonically from Δ_0 to an asymptotic value $\Delta(\tau \to \infty) := \Delta_\infty$, where Δ_∞ is determined by $\mathrm{d}V/\mathrm{d}\Delta\,|_{\Delta=\Delta_\infty} = 0$. This translates into a first condition to be imposed. A second equation comes from the energy conservation condition: $V(\Delta_0, \Delta_0) = V(\Delta_\infty, \Delta_0)$. Combined with the usual equation for the mean μ and the overlap κ, the system of equations to be solved becomes:

$$\begin{aligned} &\mu = M_m \kappa \\ &\kappa = M_n \langle [\phi_i] \rangle + \rho\kappa \left\langle \left[\phi_i'\right] \right\rangle \\ &\frac{\Delta_0^2 - \Delta_\infty^2}{2} = g^2 \left\{ \int \mathcal{D}z \Phi^2\left(\mu + \sqrt{\Delta_0}\, z\right) - \int \mathcal{D}z \left[\int \mathcal{D}x \Phi\left(\mu + \sqrt{\Delta_0 - \Delta_\infty}\, x + \sqrt{\Delta_\infty}\, z\right) \right]^2 \right\} + \Sigma_m^2 \kappa^2 (\Delta_0 - \Delta_\infty) \\ &\Delta_\infty = g^2 \int \mathcal{D}z \left[\int \mathcal{D}x \phi\left(\mu + \sqrt{\Delta_0 - \Delta_\infty}\, x + \sqrt{\Delta_\infty}\, z\right) \right]^2 + \Sigma_m^2 \kappa^2. \end{aligned} \tag{77}$$

The temporally fluctuating state is therefore described by a closed set of equations for the mean activity μ, the overlap κ, the zero-lag variance Δ_0 and the long-time variance Δ_∞. The difference $\Delta_0 - \Delta_\infty$ represents the amplitude of temporal fluctuations. If temporal fluctuations are absent, $\Delta_0 = \Delta_\infty$, and the system of equations we just derived reduces to the DMF description for stationary solutions given in Equation 40.

A similar set of equations can be derived for single unit activity. As for static stationary states, the mean activity of unit i is given by

$$\mu_i = m_i \kappa. \tag{78}$$

The static variance around this mean activity is identical for all units and given by

$$\Delta_\infty^I = g^2 \int \mathcal{D}z \left[\int \mathcal{D}x \phi\left(\mu + \sqrt{\Delta_0 - \Delta_\infty}\, x + \sqrt{\Delta_\infty}\, z\right) \right]^2 = \Delta_\infty - \Sigma_m^2 \kappa^2 \tag{79}$$

while the temporal component Δ_T^I of the variance is identical to the population averaged temporal variance

$$\Delta_T^I = \Delta_0 - \Delta_\infty. \tag{80}$$

To conclude, similarly to static stationary states, the structured connectivity P_{ij} shapes network activity in the direction defined by its right eigenvector m whenever the overlap κ does not vanish. For this reason, the mean-field theory predicts in some parameter

regions the existence of more than one chaotic solution. A formal analysis of the stability properties of the different solutions has not been performed. We nevertheless observe from numerical simulations that chaotic solutions tend to inherit the stability properties of the stationary solution they develop from. Specifically, when an homogeneous solution generates two heterogeneous bistable ones, we notice that the former loses stability in favor of the latter.

We finally observe that the critical coupling at which the DMF theory predicts the onset of chaotic fluctuations can be computed by imposing that, at the critical point, the concavity of the potential function $V(\Delta)$ is inverted (Sompolinsky et al., 1988; Harish and Hansel, 2015):

$$\left.\frac{d^2V(\Delta,\Delta_0)}{d\Delta^2}\right|_{\Delta_\infty}=0 \tag{81}$$

and the temporal component of the variance vanishes: $\Delta_0 = \Delta_\infty$. These two conditions are equivalent to the expression: $1 = g^2\langle[\phi_I'^2]\rangle$ where, as we saw, $g^2\langle[\phi_I'^2]\rangle$ coincides with the squared value of the radius of the compact component of the stability eigenspectrum (Equation 44). In the phase diagram of Figure 1B, we solved this equation for g to derive the position of the instability boundary from stationary to chaotic regimes.

Spontaneous dynamics: structures overlapping on the unitary direction

In this section, we analyze in detail a specific case, in which the connectivity vectors m and n overlap solely along the unitary direction $u = (1,1,\ldots 1)/N$. Within the statistical description of vector components, in this situation the joint probability density $p(m,n)$ can be replaced by the product two normal distributions (respectively, $\mathcal{N}(M_m,\Sigma_m^2)$ and $\mathcal{N}(M_n,\Sigma_n^2)$). The mean values M_m and M_n represent the projections of m and n on the common direction u, and the overlap between m and n is given by M_mM_n. The components m and n are otherwise independent, the fluctuations representing the remaining parts of m and n that lie along mutually orthogonal directions. In this situation, the expression for κ simplifies to

$$\begin{aligned}\kappa &= \langle n_i[\phi_i]\rangle \\ &= M_n\langle[\phi_i]\rangle\end{aligned} \tag{82}$$

so that a non-zero overlap κ can be obtained only if the mean population activity $\langle[\phi_i]\rangle$ is non-zero. Choosing independently drawn m and n vectors thus slightly simplifies the mean-field network description. The main qualitative features resulting from the interaction between the structured and the random component of the connectivity can however already be observed, and more easily understood, within this simplified setting.

Stationary solutions

The DMF description for stationary solutions reduces to a system of two non-linear equations for the population averaged mean μ and variance Δ_0:

$$\begin{aligned}\mu &= M_mM_n\langle[\phi_i]\rangle := F(\mu,\Delta_0)\\ \Delta_0 &= g^2\langle[\phi_i^2]\rangle + \Sigma_m^2M_n^2\langle[\phi_i]\rangle^2 := G(\mu,\Delta_0).\end{aligned} \tag{83}$$

The population averages $\langle[\phi_i]\rangle$ and $\langle[\phi_i^2]\rangle$ are computed as Gaussian integrals similarly to Equation 39. Equation 83 can be solved numerically for μ and Δ_0 by iterating the equations up to convergence, which is equivalent to numerically simulating the two-dimensional dynamical system given by

$$\begin{aligned}\dot{\mu}(t) &= -\mu + F(\mu,\Delta_0)\\ \dot{\Delta}_0(t) &= -\Delta_0 + G(\mu,\Delta_0),\end{aligned} \tag{84}$$

since the fixed points of this dynamical system correspond to solutions of Equation 83. Gaussian integrals in the form of $\langle[\phi_i]\rangle$ are evaluated numerically through Gauss-Hermite quadrature with a sampling over 200 points. Unstable solutions can be computed by iterating the same equations after having inverted the sign of the time variable in the first equation.

As the system of equations in Equation 83 is two-dimensional, we can investigate the number and the nature of stationary solutions through a simple graphical approach (Figure S1G). We plot on the $\mu - \Delta_0$ plane the loci of points where the two individual equations

$$\begin{aligned}\mu &= F(\mu,\Delta_0)\\ \Delta_0 &= G(\mu,\Delta_0)\end{aligned} \tag{85}$$

are satisfied. In analogy with dynamical systems approaches, we refer to the two corresponding curves as the DMF *nullclines*. The solutions of Equation 83 are then given by the intersections of the two nullclines.

To begin with, we focus on the nullcline defined by the first equation (also referred to as the μ nullcline). With respect to μ, $F(\mu,\Delta_0)$ is an odd sigmoidal function whose maximal slope depends on the value of Δ_0 and M_mM_n. When $g = 0$ and $\Sigma_m = 0$, the input variance Δ_0 vanishes. In this case, the points of the μ nullcline trivially reduce to the roots of the equation: $\mu = M_mM_n\phi(\mu)$, which admits either one ($M_mM_n < 1$), or three solutions ($M_mM_n > 1$). Non-zero values of g and Σ_m imply finite and positive values of Δ_0. As Δ_0 increases, the solutions to the equation $\mu = M_mM_n\langle[\phi_i]\rangle$ vary smoothly, delineating the full nullcline in the $\mu - \Delta_0$ plane. As in the case without

disorder ($g = 0$ and $\Sigma_m = 0$), for low structure strengths ($M_m M_n < 1$), the μ nullcline consists of a unique branch: $\mu = 0 \ \forall \Delta_0$. At high structure strengths $(M_m M_n > 1)$, instead, its shape smoothly transforms into a symmetric pitchfork.

The Δ_0 nullcline is given by the solutions of $\Delta_0 = G(\mu, \Delta_0)$ for Δ_0 as function of μ. As $G(\mu, \Delta_0)$ depends quadratically on μ, the Δ_0 nullcline has a symmetric *V*-shape centered in $\mu = 0$. The ordinate of its vertex is controlled by the parameter g, as the second term of the second equation in 83 vanishes at $\mu = 0$. For $\mu = 0$, the slope of $G(\mu, \Delta_0)$ in $\Delta_0 = 0$ is equal to g^2. As a consequence, for $g < 1$, the vertex of the Δ_0 nullcline is fixed in (0,0), while for $g > 1$, the vertex is located at $\Delta_0 > 0$ and an isolated point remains at $(0, 0)$.

The stationary solutions of the DMF equations are determined by the intersections between the two nullclines. For all values of the parameters, the nullclines intersect in $\mu = 0$, $\Delta_0 = 0$, corresponding to the trivial, homogeneous stationary solution. The existence of other solutions are determined by the qualitative features of the individual nullclines, that depend on whether $M_m M_n$ and g are smaller or greater than one (Figure S1G). The following qualitative situations can be distinguished: (i) for $M_m M_n < 1$ and $g < 1$, only the trivial solutions exist; (ii) for $M_m M_n > 1$, two additional, symmetric solutions exist for non-zero values of μ and Δ_0, corresponding to symmetric, heterogeneous stationary states; (iii) for $g > 1$, an additional solution exist for $\mu = 0$ and, corresponding to a heterogeneous solution in which individual units have non-zero stationary activity, but the population-average vanishes. For $M_m M_n > 1$, this solution can coexist with the symmetric heterogeneous ones, but in the limit of large g these solutions disappear (Figure S1G).

The next step is to assess the stability of the various solutions. As explained earlier on, the stability of the trivial state $\mu = 0$, $\Delta_0 = 0$ can be readily assessed using random matrix theory arguments (Figures S1A and S1B). This state is stable only for $M_m M_n < 1$ and $g < 1$. At $M_m M_n = 1$, it loses stability due to the outlying eigenvalue of the stability matrix, leading to the bifurcation already observed at the level of nullclines. At $g = 1$, the instability is due to the radius of the bulk of the spectrum. This leads to a chaotic state, not predicted from the nullclines for the stationary solutions.

The stability of heterogeneous stationary states is assessed by determining separately the radius of the bulk of the spectrum and the position of the outlier (Figures S1D–S1F). The radius is determined from Equation 44. The outlier is instead computed as the leading eigenvalue of the stability matrix given in Equation 71. Note that in the present framework, where the overlap is defined along the unitary direction, it is possible to show that the latter is equivalent to computing the leading stability eigenvalue of the effective dynamical system introduced in Equation 84, linearized around the corresponding fixed point. The bifurcation obtained when the outlier crosses unity is equivalent to the bifurcation predicted from the nullclines when the symmetric solutions disappear in favor of the heterogeneous solution of mean zero (Figure S1G). For $M_m M_n > 1$, we however find that as g is increased, the radius of the bulk of the spectrum always leads to a chaotic instability before the outlier becomes unstable. Correspondingly, the $\mu = 0$ and $\Delta_0 > 0$ stationary state that exist for large g is never stable.

Chaotic solutions

For large g, the instabilities of the stationary points generated by the bulk of the spectrum are expected to give rise to chaotic dynamics. We therefore turn to the DMF theory for chaotic states, which are described by an additional variable that quantifies temporal fluctuations. For the case studied here of connectivity vectors m and n overlapping only along the unitary direction, Equation 77 become

$$\mu = F(\mu, \Delta_0, \Delta_\infty) = M_m M_n \int \mathcal{D}z \phi\left(\mu + \sqrt{\Delta_0} z\right)$$

$$\Delta_0 = G(\mu, \Delta_0, \Delta_\infty) = \left[\Delta_\infty^2 + 2g^2 \left\{ \int \mathcal{D}z \Phi^2\left(\mu + \sqrt{\Delta_0} z\right) - \int \mathcal{D}z \left[\int \mathcal{D}x \Phi\left(\mu + \sqrt{\Delta_0 - \Delta_\infty} x + \sqrt{\Delta_\infty} z\right)\right]^2 \right\} + M_n^2 \Sigma_m^2 \langle [\phi_i] \rangle^2 (\Delta_0 - \Delta_\infty)\right]^{\frac{1}{2}}$$

$$\Delta_\infty = H(\mu, \Delta_0, \Delta_\infty) = g^2 \int \mathcal{D}z \left[\int \mathcal{D}x \phi\left(\mu + \sqrt{\Delta_0 - \Delta_\infty} x + \sqrt{\Delta_\infty} z\right)\right]^2 + M_n^2 \Sigma_m^2 \langle [\phi_i] \rangle^2. \quad (86)$$

As the system to be solved is now three-dimensional, graphical approaches have only limited use. Similarly to the stationary state, a practical and stable way to find numerically the solutions is to iterate the dynamical system given by

$$\begin{aligned} \dot{\mu} &= -\mu + F(\mu, \Delta_0, \Delta_\infty) \\ \dot{\Delta}_0 &= -\Delta_0 + G(\mu, \Delta_0, \Delta_\infty) \\ \dot{\Delta}_\infty &= -\Delta_\infty + H(\mu, \Delta_0, \Delta_\infty). \end{aligned} \quad (87)$$

where the double Gaussian integrals from Equation 86 can be evaluated numerically as two nested Gauss-Hermite quadratures. Note that stationary states simply correspond to solutions for which $\Delta_0 = \Delta_\infty$.

As for stationary solutions, different types of chaotic solutions appear depending on the values of the structure strength $M_m M_n$ and the disorder strength g. If $g > 1$ and $M_m M_n < 1$, a single chaotic state exists corresponding to $\mu = 0$ and $\Delta_\infty = 0$, meaning that the temporally averaged activity of all units vanishes, so that fluctuations are only temporal (Figure 1B red). As $M_m M_n$ crosses unity,

two symmetric states appear with non-zero values of μ and Δ_∞. These states correspond to bistable heterogeneous chaotic states (Figure 1B orange) that are analogous to bistable heterogeneous stationary states.

The critical disorder strength g_B at which heterogeneous chaotic states emerge (gray boundary in the phase diagram of Figure 1) is computed by evaluating the linear stability of the dynamics in 87 around the central solution $(0, \Delta_0, 0)$. A long but straightforward algebra reveals that the stability matrix, evaluated in, is simply given by

$$\begin{pmatrix} M_m M_n \langle \phi' \rangle & 0 & 0 \\ 0 & \frac{g^2\left(\langle \phi^2 \rangle + \langle \Phi \phi' \rangle - \langle \Phi \rangle \langle \phi' \rangle\right)}{\Delta_0} & 0 \\ 0 & 0 & g^2 \langle \phi' \rangle^2 \end{pmatrix}, \tag{88}$$

such that g_B corresponds to the value of the random strength g for which the largest of its three eigenvalues crosses unity.

Spontaneous dynamics: structures overlapping on an arbitrary direction

In the previous section, we focused on the simplified scenario where the connectivity vectors m and n overlapped only in the unitary direction. Here, we briefly turn to the opposite case where the overlap along the unitary direction u vanishes (i.e., $M_m = 0$, $M_n = 0$), but the overlap ρ along a direction orthogonal to u is non-zero. As we will show, although the equations describing the network activity present some formal differences, they lead to qualitatively similar regimes. The same qualitative results apply as well to the general case, where an overlap exists on both the unitary and an orthogonal direction.

The network dynamics can be studied by solving the DMF Equations 40 and 77 by setting $\mu = 0$. Stationary solutions are now determined by:

$$\begin{aligned} \kappa &= \rho \kappa \Sigma_m \Sigma_n \langle [\phi'_i(0, \Delta_0)] \rangle \; := F(\kappa, \Delta_0) \\ \Delta_0 &= g^2 \left\langle \left[\phi_i^2(0, \Delta_0)\right] \right\rangle + \Sigma_m^2 \kappa^2 \; := G(\kappa, \Delta_0). \end{aligned} \tag{89}$$

Note that, in this more general case, the relevant first-order statistics of network activity is given by the overlap κ, which now can take non-zero values even when the population-averaged activity $\langle [\phi_i] \rangle$ vanishes.

As in the previous case, the stationary solutions can be analyzed in terms of nullclines (Figure S2A). The main difference lies in the κ nullcline given by $\kappa = \rho \kappa \Sigma_m \Sigma_n \langle [\phi'_i(0, \Delta_0)] \rangle$. As both sides of the first equation are linear and homogeneous in κ, two classes of solutions exist: a trivial solution ($\kappa = 0$ for any Δ_0), and a non-trivial one ($\Delta_0 = \tilde{\Delta}_0$ for any κ), with $\tilde{\Delta}_0$ determined by:

$$\left\langle \left[\phi'_i(0, \tilde{\Delta}_0)\right] \right\rangle = 1/(\rho \Sigma_m \Sigma_n). \tag{90}$$

Because $0 < \phi'(x) < 1$, Equation 90 admits non-trivial solutions only for sufficiently large overlap values: $\rho > 1/\Sigma_m \Sigma_n$. In consequence, the κ nullcline takes qualitatively different shapes depending on the value of ρ: (i) for $\rho < 1/\Sigma_m \Sigma_n$, it consists only of a vertical branch $\kappa = 0$ (ii) for $\rho > 1/\Sigma_m \Sigma_n$ an additional horizontal branch $\Delta_0 = \tilde{\Delta}_0$ appears (Figure S2A).

The Δ_0 branch is qualitatively similar to the previously studied case of m and n overlapping along the unitary direction, with a qualitative change when the disorder parameter g crosses unity.

The stationary solutions are given by the intersections between the two nullclines. Although the shape of the κ nullcline is distinct from the shape of the μ nullcline studied in the previous case, qualitatively similar regimes are found. The trivial stationary state $\kappa = 0$, $\Delta_0 = 0$ exists for all parameter values. When the structure strength $\rho \Sigma_m \Sigma_n$ exceeds unity, two symmetric heterogeneous states appear with non-zero κ values of opposite signs (but vanishing mean μ). Finally for large g an additional state appears with $\kappa = 0$, $\Delta_0 > 0$.

Similarly to Figure 1, the solutions of Equation 89, which correspond to stationary activity states, are shown in blue in Figures S2B–S2D.

In Figure S2B we address their stability properties: again we find that when non-centered stationary solutions exist, the central fixed point becomes unstable. The instability is led by the outlier eigenvalue of the stability eigenspectrum. Similarly to Figure 1, furthermore, the DMF theory predicts an instability to chaotic phases for high g values. As for stationary states, both heterogeneous and homogeneous chaotic solutions are admitted (Figures S2C and S2D); heterogeneous chaotic states exist in a parameter region where the values of g and ρ are comparable.

Response to external inputs

In this section, we examine the effect of non-vanishing external inputs on the network dynamics. We consider the situation in which every unit receives a potentially different input I_i, so that the pattern of inputs at the network level is characterized by the N-dimensional vector $I = \{I_i\}$. The network dynamics in general depend on the geometrical arrangement of the vector I with respect to the connectivity vectors m and n. Within the statistical description used in DMF theory, the input pattern is therefore characterized by the first- and second-order statistics M_I and Σ_I of its elements, as well as by the value of the correlations Σ_{mI} and Σ_{nI} with the vectors m and n. In geometric terms, M_I quantifies the component of I along the unit direction u, while Σ_{mI} and Σ_{nI} quantify the overlaps with m

and n along directions orthogonal to u. For the sake of simplicity, here we consider two connectivity vectors m and n that overlap solely on the unitary direction ($\rho = 0$). The two vectors thus read (see Equation 34):

$$\begin{aligned} m &= M_m + \Sigma_m x_1 \\ n &= M_n + \Sigma_n x_2. \end{aligned} \tag{91}$$

The input pattern can overlap with the connectivity vectors on the common (u) and on the orthogonal directions (x_1 and x_2). It can moreover include further orthogonal components of strength $\Sigma_\perp$. The most general expression for the input vector can thus be written as:

$$I = M_I + \frac{\Sigma_{mI}}{\Sigma_m} x_1 + \frac{\Sigma_{nI}}{\Sigma_n} x_2 + \Sigma_\perp h \tag{92}$$

where h is a standard normal vector. We first focus on the equilibrium response to constant inputs, and then turn to transient dynamics.

The mean-field equations in presence of external inputs can be derived in a straightforward fashion by following the same steps as in the input-free case. We start by considering the statistics of the effective coupling term, which is given by $\xi_i(t) = \eta_i(t) + I_i(t)$, with $\eta_i(t)$ defined as in Equation 20. We can then exploit the statistics of $\eta_i(t)$ which have been computed in the previous paragraphs to obtain the equation for the mean activity:

$$\mu_i = [x_i] = m_i \kappa + I_i. \tag{93}$$

Equation 93 indicates that the direction of the average network activity is determined by a combination of the structured recurrent connectivity and the external input pattern. The final direction of the activation vector in the N-dimensional population space is controlled by the value of the overlap κ, which depends on the relative orientations of m, n and I. Its value is given by the self-consistent equation:

$$\begin{aligned} \kappa &= \langle n_i [\phi_i] \rangle \\ &= \left\langle n_i \int \mathcal{D}z \phi \left(m_i \kappa + I_i + \sqrt{\Delta_0^I} z \right) \right\rangle \\ &= M_n \langle [\phi_i] \rangle + \Sigma_{nI} \langle [\phi_i'] \rangle, \end{aligned} \tag{94}$$

as both vectors m and I share non-trivial overlap directions with n.

The second-order statistics of the noise are given by:

$$[\xi_i(t)\xi_j(t+\tau)] = \delta_{ij} g^2 \langle [\phi_i(t)\phi_i(t+\tau)] \rangle + m_i m_j \kappa^2 + (m_i I_j + m_j I_i)\kappa + I_i I_j. \tag{95}$$

Averaging across the population we obtain:

$$\langle [\xi_i(t)\xi_i(t+\tau)] \rangle - \langle [\xi_i(t)] \rangle^2 = g^2 \langle [\phi_i^2] \rangle + \Sigma_m^2 \kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2. \tag{96}$$

The first term of the r.h.s. represents the quenched variability inherited from the random connectivity matrix, while $\Sigma_\mu^2 = \Sigma_m^2 \kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2$ represents the variance induced by the structure, which is inherited from both vectors m and I (Equation 93). From Equation 92, the variance of the input reads:

$$\Sigma_I^2 = \frac{\Sigma_{mI}^2}{\Sigma_m^2} + \frac{\Sigma_{nI}^2}{\Sigma_n^2} + \Sigma_\perp^2. \tag{97}$$

The final DMF equations to be solved are given by the following system:

$$\begin{aligned} \mu &= M_m \kappa + M_I \\ \ddot{\Delta} &= \Delta - \left\{ g^2 \langle [\phi_i(t)\phi(t+\tau)] \rangle + \Sigma_m^2 \kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2 \right\} \\ \kappa &= M_n \langle [\phi_i] \rangle + \Sigma_{nI} \langle [\phi_i'] \rangle \end{aligned} \tag{98}$$

which, similarly to the cases we examined in detail so far, admits both stationary and chaotic solutions. As for spontaneous dynamics, the instabilities to chaos are computed by evaluating the radius of the eigenspectrum of the stability matrix S_{ij} (Equation 44). The stability matrix can admit an outlier eigenvalue as well, whose value can be predicted with a mean-field stability analysis. Extending the arguments already presented in the previous paragraphs allows to show that the effective stability matrix $\mathcal{M}$ is given by:

$$\mathcal{M} = \begin{pmatrix} 0 & 0 & M_m \\ 2g^2 \langle [\phi_i \phi_i'] \rangle & g^2 \{ \langle [\phi_i'^2] \rangle + \langle [\phi_i \phi_i''] \rangle \} & 2\Sigma_m^2 \kappa^0 + 2\Sigma_{mI} \\ 2bg^2 \langle [\phi_i \phi_i'] \rangle & bg^2 \{ \langle [\phi_i'^2] \rangle + \langle [\phi_i \phi_i''] \rangle \} & b(2\Sigma_m^2 \kappa^0 + 2\Sigma_{mI}) + a \end{pmatrix}, \tag{99}$$

with:

$$a = M_m M_n \langle [\phi_i'] \rangle + M_m \Sigma_{nl} \langle [\phi_i''] \rangle$$
$$b = \frac{1}{2} \{ M_n \langle [\phi_i''] \rangle + \Sigma_{nl} \langle [\phi_i'''] \rangle \}. \quad (100)$$

As in the input-free case, when the stability eigenspectrum contains one outlier eigenvalue, its position is well predicted by the largest eigenvalue of $\mathcal{M}$.

In the following, we refer to Figure 2 and analyze in detail the contribution of every input direction to the final network dynamics.

In Figure 2D (left), we consider a unit-rank structure whose vectors m and n are orthogonal: $M_m = M_n = 0$. The input direction is orthogonal to the connectivity vectors: $\Sigma_{ml} = \Sigma_{nl} = 0$, so that the input strength is quantified by the amplitude of the component along h ($\Sigma_\perp$). In this configuration, because of Equation 94, the amount of structured activity quantified by κ systematically vanishes.

In Figure 2D (center), we consider again orthogonal connectivity vectors, but we take an input pattern which overlaps with n along x_2. We keep $\Sigma_\perp = 1$ fixed and we vary the component of the input along n by increasing Σ_{nl}. As can be seen from the equation for κ (Equation 98), the overlap Σ_{nl} between the input and the left vector n has the effect of increasing the value of κ, which would otherwise vanish since the structure has null strength ($M_n = 0$). In response to the input, a structured state emerges. From the same equation, furthermore, one can notice that the Σ_{nl} term has the effect of breaking the sign reversal symmetry ($x \to -x$) that characterizes the mean-field equations in the case of spontaneous dynamics.

In Figure 2D (right), we include strong non-vanishing structure strengths ($M_m M_n = 3.5$). In absence of external activity, the network dynamics thus admit two bistable solutions (Figure 1). We consider an input pattern that correlates with n but is orthogonal to the structure overlap direction ($M_I = 0$, $\Sigma_{nl} > 0$). In this configuration, the external input has the effect of disrupting the symmetry between the two stable solutions. For sufficiently strong input values, one of the two stable solutions disappears by annihilating with the unstable one.

In Figure S4C, we show that the value of the critical input strength for which one of the two stable solution disappears can be controlled by an additional external input that overlaps with n on a different, orthogonal direction. Specifically, in Figure S4C, we tune the additional input along the direction of the structure overlap u. This input component can be thought as a modulatory signal which controls the way the network dynamics process the input stimulus along x_2. In models of computational tasks that employ non-linear input responses (Figure 4), a modulatory input along the structure overlap can regulate the threshold value of the input strength that the network has learnt to detect. Similarly, in Figures 5 and 6, modulatory inputs are used to completely block the response to the non-relevant input stimulus, so that the readout can produce context-dependent outputs.

Asymmetric solutions

A major effect of external inputs is that they break the sign reversal symmetry ($x \to -x$) present in the network dynamics without inputs. As a consequence, in the parameter regions where the network dynamics admit bistable structured states, the two stable solutions are characterized by different statistics and stability properties.

To illustrate this effect, we focus on the simple case where the external input pattern I overlaps with the connectivity vectors m and n solely on the unitary direction ($M_I \neq 0$, $\Sigma_{ml} = \Sigma_{nl} = 0$). The solutions of the system of equations corresponding to stationary states can be visualized with the help of the graphical approach, which unveils the symmetry breaking of network dynamics induced by external inputs (Figure S4D).

Similarly to the input-free case, the Δ_0 nullcline consists of a symmetric V-shaped curve. In contrast to before, however, the vertex of the nullcline is no longer fixed in $(0, 0)$, but takes positive ordinate values also at low g values. The value of $G(0, \Delta_0)$, indeed, does not vanish, because of the finite contribution from the input pattern Σ_I^2.

The nullcline curves of μ are instead strongly asymmetric. For low $M_m M_n$ values, one single μ nullcline exists. In contrast to the input-free case, this nullcline is no longer centered in zero. As a consequence, it intersects the Δ_0 nullclines in one non-zero point, corresponding to a unique heterogeneous stationary solution. As $M_m M_n$ increases, a second, separated branch can appear. In contrast to the input-free case, the structure strength at which the second branch appears is not always equal to unity, but depends on the mean value of the input. If $M_m M_n$ is strong enough, the negative branch of the nullcline can intersect the Δ_0 nullcline in two different fixed points, while a third solution is built on the positive μ nullcline. As g increases, the two intersections on the negative branch become closer and closer and they eventually collapse together. At a critical value g_B, the network activity discontinuously jumps from negative to positive mean solutions.

As they are no longer symmetrical, the stability of the positive and the negative fixed points has to be assessed separately, and gives rise to different instability boundaries. Computing the position of the outlier reveals that, when more than one solution is admitted by the mean-field system of equations, the centered one is always unstable.

As the stability boundaries of different stationary solutions do not necessarily coincide, in presence of external input patterns the phase diagram of the dynamics are in general more complex (Figures S4A–S4C). Specifically, hybrid dynamical regimes, where one static solution co-exists with a chaotic attractor, can be observed.

Transient dynamics

We now turn to transient dynamics evoked by a temporal step in the external input (Figure 2B). We specifically examine the projection of the activation vector and its average onto the two salient directions spanned by vectors m and I.

The transient dynamics of relaxation to a stationary solution can be assessed by linearizing the mean-field dynamics. We compute the time course of the average activation vector μ_i, and we finally project it onto the two orthogonal directions which are indicated in the small insets of Figure 2B.

Similarly to Equation 45, the time evolution of μ_i is governed by:

$$\dot{\mu}_i(t) = -\mu_i(t) + m_i\kappa(t) + I_i(t) \tag{101}$$

so that, at every point in time:

$$\mu_i(t) = m_i\tilde{\kappa}(t) + \tilde{I}_i(t), \tag{102}$$

where $\tilde{\kappa}(t)$ and $\tilde{I}_i(t)$ coincide with the low-pass filtered versions of $\kappa(t)$ and $I(t)$.

When the network activity is freely decaying back to an equilibrium stationary state, $\tilde{I}_i(t)$ coincides with a simple exponential relaxation to the pattern I_i. The decay timescale is set by the time evolution of activity (Equation 6), which is taken here to be equal to unity:

$$\tilde{I}_i(t) = I_i + \left(I_i^{ic} - I_i\right)e^{-t}. \tag{103}$$

The timescale of $\tilde{\kappa}(t)$ is inherited from the dynamics of $\kappa(t)$. We thus refer to our mean-field stability analysis, and we compute the relaxation time of the population statistics $\kappa(t)$ as the largest eigenvalue of the stability matrix $\mathcal{M}$. The eigenvalue predicts a time constant τ_r, which is in general larger than unity. As a consequence, the relaxation of $\kappa(t)$ obeys, for small displacements:

$$\kappa(t) = \kappa^0 + \left(\kappa^{ic} - \kappa^0\right)e^{-\frac{t}{\tau_r}}, \tag{104}$$

where the asymptotic value of κ^0 is determined from the equilibrium mean-field equations (Equations 98). Finally, the time course of $\tilde{\kappa}(t)$ is derived as the low-pass filter version of Equation 104 with unit decay timescale.

Rank-two connectivity structures

In the following paragraphs, we provide the detailed analysis for network models with rank-two connectivity structures. The structured component of the connectivity can be written as:

$$P_{ij} = \frac{m_i^{(1)}n_j^{(1)}}{N} + \frac{m_i^{(2)}n_j^{(2)}}{N}, \tag{105}$$

where the vector pairs $m^{(1)}$ and $m^{(2)}$, $n^{(1)}$ and $n^{(2)}$ are assumed to be linearly independent.

As in the case of unit-rank structures, we determine the network statistics by exploiting the link between linear stability analysis and mean-field description. The study of the properties of eigenvalues and eigenvectors for the low-dimensional matrix P_{ij} helps to predict the complex behavior of activity above the instability and to restrict our attention to the cases of interest.

The mean activity of the network in response to a fixed input pattern I_i is given by:

$$\mu_i = \kappa_1 m_i^{(1)} + \kappa_2 m_i^{(2)} + I_i. \tag{106}$$

The final direction of the population activity is thus determined by the overlap values $\kappa_1 = \langle n_i^{(1)}[\phi_i]\rangle$ and $\kappa_2 = \langle n_i^{(2)}[\phi_i]\rangle$.

The expression of the mean-field equations for the first- and second-order statistics are determined by the geometrical arrangement of the connectivity and the input vectors. Similarly to the unit-rank case, the simplest mean-field solutions correspond to stationary states, which inherit the structure of the most unstable eigenvectors of the connectivity matrix J_{ij}. The stability of the heterogeneous stationary states can be assessed as before by evaluating separately the value of the radius (Equation 44) and the position of the outliers of the linear stability matrix S_{ij}.

Similarly to the unit-rank case, it is possible to compute the position of the outlier eigenvalues by studying the linearized dynamics of the network statistics close to the fixed point, that is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\mu^1\\ \Delta_0^1\\ \kappa_1^1\\ \kappa_2^1\end{pmatrix} = -\begin{pmatrix}\mu^1\\ \Delta_0^1\\ \kappa_1^1\\ \kappa_2^1\end{pmatrix} + \mathcal{M}\begin{pmatrix}\mu^1\\ \Delta_0^1\\ \kappa_1^1\\ \kappa_2^1\end{pmatrix}. \tag{107}$$

Note that, in κ_k^l, the subscript $k = 1, 2$ refers to the left vector $n^{(k)}$ with which the overlap is computed, while the superscript $l = 0, 1$ indicates the order of the perturbation away from the fixed point.

In order to compute the elements of the linear stability matrix $\mathcal{M}$, we follow and extend the reasoning discussed in details for the unit-rank case. We start by considering the time evolution of the linearized activity μ_i^1, which similarly to Equation 45 reads:

$$\dot{\mu}_i^1(t) = -\mu_i^1 + m_i^{(1)}\kappa_1^1 + m_i^{(2)}\kappa_2^1. \tag{108}$$

At every point in time, we can write: $\mu_i^t = m_i^{(1)}\tilde{\kappa}_1^t + m_i^{(2)}\tilde{\kappa}_2^t$, where $\tilde{\kappa}_k^t$ is the low-pass filtered version of κ_k^t: $(1 + \mathrm{d}/\mathrm{d}t)\tilde{\kappa}_k^t = \kappa_k^t$.
In the case of orthogonal (zero mean), random connectivity vectors, we get:

$$\dot{\mu}^1(t) = -\mu^1, \tag{109}$$

so that the elements in the first row of $\mathcal{M}$ vanish. In analogy with Equation 64, the linearized dynamics of Δ_0 gives instead:

$$\dot{\Delta}_0^1 = -\Delta_0^1 + 2g^2\left\langle\left[\phi_i\phi_i'\right]\right\rangle\mu^1 + g^2\left\{\left\langle\left[\phi_i'^2\right]\right\rangle + \left\langle\left[\phi_i\phi_i''\right]\right\rangle\right\}\Delta_0^1 + 2\Sigma_m^2\kappa_1^0\kappa_1^1 + 2\Sigma_m^2\kappa_2^0\kappa_2^1. \tag{110}$$

Similarly to the unit-rank case (Equation 47), in order to determine the linear response of κ_1 we need to compute:

$$\kappa_1^1 = \left\langle n_i^{(1)}\left[x_i^1\phi'\left(x_i^0\right)\right]\right\rangle = \left\langle n_i^{(1)}\mu_i\left[\phi_i'\right]\right\rangle + \left(\frac{\Delta_0^1}{2} - \left\langle\mu_i^1\mu_i^0\right\rangle - \left\langle\mu_i^1\right\rangle\left\langle\mu_i^0\right\rangle\right)\left\langle n_i^{(1)}\left[\phi_i''\right]\right\rangle \tag{111}$$

A similar expression can be derived for κ_2^1.
In general, the integrals in the r.h.s. can be expressed in terms of the perturbations $\tilde{\kappa}_1^1$, $\tilde{\kappa}_2^1$ and Δ_0^1, leading to expressions of the form:

$$\begin{aligned}\kappa_1^1 &= a_{11}\tilde{\kappa}_1^1 + a_{12}\tilde{\kappa}_2^1 + b_1\Delta_0^1\\ \kappa_2^1 &= a_{21}\tilde{\kappa}_1^1 + a_{22}\tilde{\kappa}_2^1 + b_2\Delta_0^1.\end{aligned} \tag{112}$$

Applying the operator $(1 + \mathrm{d}/\mathrm{d}t)$ to the Equation 111 allows to reshape the results in the final matrix form:

$$\mathcal{M} = \begin{pmatrix} 0 & 0 & 0 & 0\\ 2g^2\left\langle\left[\phi_i\phi_i'\right]\right\rangle & g^2\left\{\left\langle\left[\phi_i'^2\right]\right\rangle + \left\langle\left[\phi_i\phi_i''\right]\right\rangle\right\} & 2\Sigma_m^2\kappa_1^0 & 2\Sigma_m^2\kappa_2^0\\ 2b_1g^2\left\langle\left[\phi_i\phi_i'\right]\right\rangle & b_1g^2\left\{\left\langle\left[\phi_i'^2\right]\right\rangle + \left\langle\left[\phi_i\phi_i''\right]\right\rangle\right\} & 2b_1\Sigma_m^2\kappa_1^0 + a_{11} & 2b_1\Sigma_m^2\kappa_2^0 + a_{12}\\ 2b_2g^2\left\langle\left[\phi_i\phi_i'\right]\right\rangle & b_2g^2\left\{\left\langle\left[\phi_i'^2\right]\right\rangle + \left\langle\left[\phi_i\phi_i''\right]\right\rangle\right\} & 2b_2\Sigma_m^2\kappa_1^0 + a_{21} & 2b_2\Sigma_m^2\kappa_2^0 + a_{22}\end{pmatrix}, \tag{113}$$

where the values of the constants *a* and *b* depend on the geometric arrangement of the structure and the input vectors.

In the following, we consider several specific cases of interest. Note that the non-linear network dynamics is determined by the relative orientation of the structure and input vectors, but also by the characteristics of the statistical distribution of their elements. In contrast to the cases we analyzed so far, the precise shape of the distribution of the entries in the connectivity vectors can play an important role when the rank of P_{ij} is larger than unity. In the following, we focus on the case of broadly, normally distributed patterns.

Rank-two structures with null overlap

The simplest case we consider consists of rank-two matrices whose four connectivity vectors $m^{(1)}$, $m^{(2)}$, $n^{(1)}$, and $n^{(2)}$ are mutually orthogonal. From the point of view of responses to inputs, networks with this structure behave as superpositions of two independent unit-rank structures.

Similarly to the unit-rank case, if the connectivity vectors are orthogonal, the network is silent in absence of external inputs: $\kappa^1 = \kappa^2 = 0$. A single homogeneous state – stationary or chaotic – is the unique stable attractor of the dynamics. Consistently, the eigenspectrum of J_{ij} does not contain any outlier, since every eigenvalue of P_{ij} vanishes.

In order to compute the eigenspectrum of P_{ij}, we can rotate the matrix onto a basis defined by an orthonormal set of vectors, and compute its eigenvalues in the transformed basis. For simplicity, we consider an orthonormal set whose first four vectors are built from the connectivity vectors:

$$\begin{aligned}u_1 &= \alpha_1 m^{(1)}\\ u_2 &= \alpha_2 m^{(2)}\\ u_3 &= \alpha_3 n^{(1)}\\ u_4 &= \alpha_4 n^{(2)},\end{aligned} \tag{114}$$

where the coefficient α_k $(k = 1, \ldots, 4)$ denote the normalization factors. In this basis, the first four rows and columns of the rotated matrix P_{ij}' read:

$$P_{ij}' = \frac{1}{N}\begin{pmatrix} 0 & 0 & \frac{1}{\alpha_1\alpha_3} & 0\\ 0 & 0 & 0 & \frac{1}{\alpha_2\alpha_4}\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \tag{115}$$

all the remaining entries being fixed to 0. From the present matrix form, it easy to verify that all the eigenvalues of P_{ij}', and thus all the eigenvalues of P_{ij}, vanish. Note that rewriting P_{ij} in an orthonormal basis simplifies the search for its eigenvalues also in more complex cases where the connectivity vectors share several overlap directions. In those cases, a proper basis needs to be built starting from the connectivity vectors through a Gram-Schmidt orthonormalization process.

As a side note we observe that, even though P'_{ij} (and thus P_{ij}) admits only vanishing eigenvalues, its rank is still equal to two. Indeed, the rank can be computed as N minus the dimensionality of the kernel associated to P'_{ij}, defined by any vector x obeying $P'x = 0$. As P'_{ij} contains $N-2$ empty rows, the last equations impose two independent contraints on the components of x. As a consequence, the dimensionality of the kernel equals $N-2$, and the rank is equal to two.

We turn to responses that are obtained in presence of external inputs. We examine the network dynamics in response to a normalized input I which partially correlates with one of the left-connectivity vectors, here $n^{(1)}$:

$$I = n^{(1)}\frac{\Sigma_{nI}}{\Sigma_n^2} + x\sqrt{\Sigma_I^2 - \frac{\Sigma_{nI}^2}{\Sigma_n^4}}. \tag{116}$$

Similarly to the unit-rank case, we find that I elicits a network response in the plane $I - m^{(1)}$. The overlap values are given by:

$$\begin{aligned}\kappa_1 &= \Sigma_{nI}\langle[\phi_i']\rangle\\ \kappa_2 &= 0,\end{aligned} \tag{117}$$

and they can be used to close the mean-field equations together with the equation for the first ($\mu = 0$) and second-order statistics. In the case of stationary states we have:

$$\Delta_0 = g^2\langle[\phi_i^2]\rangle + \Sigma_m^2(\kappa_1^2 + \kappa_2^2) + \Sigma_I^2. \tag{118}$$

Similar arguments allow to derive the two equations needed for the chaotic states.

In order to assess the stability of the stationary states, we evaluate the position of the outliers in the stability eigenspectrum by computing the eigenvalues of $\mathcal{M}$ (Equation 113). In the case of orthogonal structures and correlated input patterns I, a little algebra reveals that all the a values vanish, while we have:

$$\begin{aligned}b^1 &= \frac{1}{2}\Sigma_{nI}\langle[\phi_i'']\rangle\\ b^2 &= 0.\end{aligned} \tag{119}$$

We conclude that the first and the last row of $\mathcal{M}$ always vanish. Furthermore, the second and the third rows are proportional one to the other. As a consequence, the stability analysis predicts at most one outlier eigenvalue, which is indeed observed in the spectrum (not shown). The outlier is negative, as the effect of introducing inputs in the direction of the left vector $n^{(1)}$ is to further stabilize the dynamics. As it will be shown, more than one outlier can be observed in the case where the low-dimensional structure involves overlap directions.

Rank-two structures with internal pairwise overlap

As a second case, we consider structured matrices where the two connectivity pairs $m^{(1)}$ and $n^{(1)}$, $m^{(2)}$ and $n^{(2)}$ share two different overlap directions, defined by vectors y_1 and y_2. We set:

$$\begin{aligned}m^{(1)} &= \sqrt{\Sigma^2 - \rho_1^2}\,x_1 + \rho_1 y_1\\ m^{(2)} &= \sqrt{\Sigma^2 - \rho_2^2}\,x_2 + \rho_2 y_2\\ n^{(1)} &= \sqrt{\Sigma^2 - \rho_1^2}\,x_3 + \rho_1 y_1\\ n^{(2)} &= \sqrt{\Sigma^2 - \rho_2^2}\,x_4 + \rho_2 y_2.\end{aligned} \tag{120}$$

where Σ^2 is the variance of the connectivity vectors and ρ_1^2 and ρ_2^2 quantify the overlaps along the directions y_1 and y_2.

By rotating P_{ij} onto the orthonormal basis that can be built from $m^{(1)}$ and $m^{(2)}$ by orthogonalizing the left vectors $n^{(1)}$ and $n^{(2)}$, one can easily check that the two non-zero eigenvalues of P_{ij} are given by $\lambda_1 = \rho_1^2$ and $\lambda_2 = \rho_2^2$. They correspond, respectively, to the two right-eigenvectors $m^{(1)}$ and $m^{(2)}$. In absence of external inputs, an instability is thus likely to occur in the direction of the $m^{(k)}$ vector which corresponds to the strongest overlap.

We specifically focus on the degenerate condition where the two overlaps are equally strong, $\rho_1 = \rho_2 = \rho$, and any combination of $m^{(1)}$ and $m^{(2)}$ is a right-eigenvector. The mean-field equations for the first-order statistics read:

$$\begin{aligned}\kappa_1 &= \rho^2\kappa_1\langle[\phi_i']\rangle\\ \kappa_2 &= \rho^2\kappa_2\langle[\phi_i']\rangle.\end{aligned} \tag{121}$$

Similarly to Equation 89, the two equations admit a silent ($\kappa_1 = \kappa_2 = 0$) and a non-trivial state, determined by two identical conditions which read:

$$1 = \rho^2\langle[\phi_i'(0, \Delta_0)]\rangle. \tag{122}$$

The equation above determines the value of Δ_0. Note that the non-trivial state exists only for $\rho > 1$.

A second condition is imposed by the equation for the second-order momentum which reads, for stationary solutions:

$$\Delta_0 = g^2 \left\langle \left[\phi_i^2\right] \right\rangle + \Sigma^2 \left(\kappa_1^2 + \kappa_2^2\right). \tag{123}$$

As the value of Δ_0 is fixed, the mean-field set of equations fixes only the sum $\kappa_1^2 + \kappa_2^2$, but not each single component. The mean-field thus returns a one-dimensional continuum of solutions, the shape of which resembles a ring of radius $\sqrt{\kappa_1^2 + \kappa_2^2}$ in the $m^{(1)} - m^{(2)}$ plane (see Figures S5D and S5E). Similarly to the unit-rank case, the value of the radius can be computed explicitly by solving numerically the two mean-field equations (three in the case of chaotic regimes), and depends on the relative magnitude of ρ^2 compared to g (Figure S5F). Highly disordered connectivities have the usual effect of suppressing non-trivial structured solutions in favor of homogeneous and unstructured states. For sufficiently high g values, furthermore, structured solution can display chaotic dynamics (Figures S5E and S5F, red).

A linear stability analysis reveals that the one-dimensional solution consists of a continuous set of marginally stable states. Similarly to the orthogonal vectors case, the position of the outliers in the eigenspectra of S_{ij} can be evaluated by computing the reduced stability matrix $\mathcal{M}$, which reads:

$$\mathcal{M} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 2g^2 \left\langle \left[\phi_i \phi_i'\right] \right\rangle & g^2 \left\{ \left\langle \left[\phi_i'^2\right] \right\rangle + \left\langle \left[\phi_i \phi_i''\right] \right\rangle \right\} & 2\Sigma_m^2 \kappa_1^0 & 2\Sigma_m^2 \kappa_2^0 \\ 2b_1 g^2 \left\langle \left[\phi_i \phi_i'\right] \right\rangle & b_1 g^2 \left\{ \left\langle \left[\phi_i'^2\right] \right\rangle + \left\langle \left[\phi_i \phi_i''\right] \right\rangle \right\} & 2b_1 \Sigma_m^2 \kappa_1^0 + a_{11} & 2b_1 \Sigma_m^2 \kappa_2^0 \\ 2b_2 g^2 \left\langle \left[\phi_i \phi_i'\right] \right\rangle & b_2 g^2 \left\{ \left\langle \left[\phi_i'^2\right] \right\rangle + \left\langle \left[\phi_i \phi_i''\right] \right\rangle \right\} & 2b_2 \Sigma_m^2 \kappa_1^0 & 2b_2 \Sigma_m^2 \kappa_2^0 + a_{22} \end{pmatrix}, \tag{124}$$

with:

$$\begin{aligned} a_{11} &= \rho^2 \left\langle \left[\phi_i'\right] \right\rangle \\ b_1 &= \frac{1}{2} \rho^2 \kappa_1^0 \left\langle \left[\phi_i'''\right] \right\rangle \end{aligned} \tag{125}$$

and

$$\begin{aligned} a_{22} &= \rho^2 \left\langle \left[\phi_i'\right] \right\rangle \\ b_2 &= \frac{1}{2} \rho^2 \kappa_2^0 \left\langle \left[\phi_i'''\right] \right\rangle. \end{aligned} \tag{126}$$

As shown in Figure S5G, diagonalizing the stability matrix $\mathcal{M}$ returns the values of two distinct outlier eigenvalues. The third non-zero eigenvalue of $\mathcal{M}$ lies instead systematically inside the compact component of the spectrum, and corresponds to an average measure of the timescales inherited by the random modes. One of the two outliers is tuned exactly to the stability boundary for every value of the parameters which generate a ring solution. This marginally stable eigenvalue is responsible for the slow dynamical timescales which are observed in numerical simulations of the network activity (Figures S5D and S5E).

The DMF predictions formally hold in the limit of infinite-size networks; in simulations of finite-size networks, the dynamics instead always converge on a small number of equilibrium spontaneous states located on the ring (see Figures S5D and S5E). The equilibrium reached in a given situation is determined by the corresponding realization of the random part of the connectivity, and the initial conditions. Different realizations of the random connectivity lead to different equilibrium states, which all however lie on the predicted ring (see Figures S5D and S5E). For a given realization of the random connectivity, transient dynamics moreover show a clear signature of the ring structure. Indeed the points on the ring are close to stable and form a slow manifold. The convergence to the equilibrium activity is therefore very slow, and the temporal dynamics explore the ring structure.

We next examine how the structured, ring-shaped solution is perturbed by the injection of external input patterns.

We consider an input pattern I of variance Σ_I^2. When I does not share any overlap direction with the left vectors $n^{(1)}$ and $n^{(2)}$, the mean-field equations are affected solely by an extra term Σ_I which needs to be included in the equation for the second-order statistics (Equation 123). As the equations for the first-order statistics do not change, the one-dimensional degeneracy of the solution persists. The extra term Σ_I^2 however decreases the value of the radius of the ring.

When the input contains a component which overlaps with one or both left vectors $n^{(1)}$ and $n^{(2)}$, the degeneracy in the two equations for κ_1 and κ_2 is broken. As a consequence, the one-dimensional solution collapses onto a unique stable point. Consider for example an input pattern of the form:

$$I = \Sigma_I \left(\sqrt{1-\alpha}\, x_3 + \sqrt{\alpha}\, x_4 \right). \tag{127}$$

The equations for the first order become:

$$\begin{aligned} \kappa_1 &= \left(\rho^2 \kappa_1 + \Sigma_I \sqrt{1-\alpha} \sqrt{\Sigma^2 - \rho^2} \right) \left\langle \left[\phi_i'\right] \right\rangle \\ \kappa_2 &= \left(\rho^2 \kappa_2 + \Sigma_I \sqrt{\alpha} \sqrt{\Sigma^2 - \rho^2} \right) \left\langle \left[\phi_i'\right] \right\rangle \end{aligned} \tag{128}$$

or, alternatively:

$$\kappa_1 = \frac{\Sigma_I \sqrt{1-\alpha}\sqrt{\Sigma^2 - \rho^2}\langle [\phi_i'] \rangle}{1 - \rho^2 \langle [\phi_i'] \rangle}$$
$$\kappa_2 = \frac{\Sigma_I \sqrt{\alpha}\sqrt{\Sigma^2 - \rho^2}\langle [\phi_i'] \rangle}{1 - \rho^2 \langle [\phi_i'] \rangle}. \tag{129}$$

The values of κ_1 and κ_2 are thus uniquely specified, and can be computed by iterating the two equations together with the expression for the second-order statistics:

$$\Delta_0 = g^2 \langle [\phi_i^2] \rangle + \Sigma^2 (\kappa_1^2 + \kappa_2^2) + \Sigma_I^2. \tag{130}$$

In a similar way, the presence of correlated external inputs affect the values of the entries of the reduced stability matrix $\mathcal{M}$:

$$b_1 = \frac{1}{2}\left(\rho^2 \kappa_1^0 + \Sigma_I \sqrt{1-\alpha}\sqrt{\Sigma^2 - \rho^2}\right)\langle [\phi_i'''] \rangle$$
$$b_2 = \frac{1}{2}\left(\rho^2 \kappa_2^0 + \Sigma_I \sqrt{\alpha}\sqrt{\Sigma^2 - \rho^2}\right)\langle [\phi_i'''] \rangle. \tag{131}$$

In Figures S5H and S5I, we focus on the case of an external input pattern aligned with x_3 (and thus $n^{(1)}$). We fix $\alpha = 0$, that implies $\kappa_2 = 0$.

Solving the mean-field equations reveals that, according to the strength of the input Σ_I, one or three fixed points exist. When the input is weak with respect to the structure overlap ρ^2, two fixed points appear in the proximity of the ring, along the direction defined by the axis $\kappa^2 = 0$ (Figure S5H, top). In particular, when I positively correlates with $n^{(1)}$, only the fixed point with positive value of κ_1 gets stabilized. The remaining two solutions are characterized by one outlier eigenvalue which lays above the instability boundary, and are thus unstable. On the other hand, when the input is sufficiently strong, solely the stable fixed point survives (Figure S5H, bottom). Activity is then robustly projected in the direction defined by the right vector $m^{(1)}$.

Rank-two structures for oscillations

We finally consider the following configuration:

$$\begin{aligned} m^{(1)} &= \alpha x_1 + \rho y_1 \\ m^{(2)} &= \alpha x_2 + \rho y_2 \\ n^{(1)} &= \alpha x_3 + \rho y_2 + \gamma \rho y_1 \\ n^{(2)} &= \alpha x_4 - \rho y_1, \end{aligned} \tag{132}$$

where the right- and the left-connectivity vectors share two cross-overlap directions y_1 and y_2. Note that the vectors in one of the two pairs, $m^{(1)} - n^{(2)}$, are negatively correlated. A second overlap is introduced internally to the $m^{(1)} - n^{(1)}$ pair, and scales with the parameter γ. The directions x_j, with $k = 1, \ldots, 4$, represent uncorrelated terms. Note that different values of α affect quantitatively the network statistics, but they do not change the phase diagram in Figure S8A.

By rotating P_{ij} on a proper orthonormal basis, one can check that its eigenvalues are given by:

$$\lambda_{\pm} = \frac{\gamma \rho^2}{2}\left(1 \pm \sqrt{1 - \frac{4}{\gamma^2}}\right), \tag{133}$$

and they are complex conjugate for $\gamma < 2$. In this case, the internal overlap γ has the effect of returning a non-vanishing real part. The two complex conjugate eigenvectors are given by:

$$e^{\pm} = \left(-\frac{\gamma}{2} m^{(1)} + m^{(2)}\right) \pm i \sqrt{\left|1 - \frac{4}{\gamma^2}\right|}\, m^{(1)}. \tag{134}$$

The eigenspectrum of $J_{ij} = g\chi_{ij} + P_{ij}$ inherits the pair of non-zero eigenvalues of P_{ij}. When $g < 1$ and $\gamma < 2$, the trivial fixed point thus undergoes a Hopf bifurcation when the real part of λ crosses unity (Figure S8A, blue). When $\gamma > 2$, instead, the two eigenvalues are real. One bifurcation to bistable stationary activity occurs when the largest eigenvalue λ_+ crosses unity (Figure S8A, gray).

On the boundary corresponding to the Hopf bifurcation, the frequency of instability ω_H is determined by the imaginary part of Equation 133. At the instability, the oscillatory activity of unit i can be represented as a point on the complex plane. Since close to the bifurcation we can write:

$$\mu_i = e_i^+ e^{i\omega_H t} + c.c., \tag{135}$$

its coordinates are given by the real and the imaginary part of the ith component of the complex eigenvector e^+. The phase of oscillation can then be computed as the angle defined by this point with respect to the real axis. Note that the disorder in the elements of the eigenvector, which is inherited by the random distribution of the entries of the connectivity vectors $m^{(1)}$ and $m^{(2)}$, tends to favor a broad distribution of phases across the population.

In the limit case where the real and the imaginary parts of the complex amplitude of the oscillators are randomly and independently distributed, the population response resembles a circular cloud in the complex plane. In this case, the phase distribution across the population is flat. Note that a completely flat phase distribution can be obtained for arbitrary frequency values by adopting a rank-two structure where an internal overlap of magnitude $\gamma\rho^2$ exists between vectors $m^{(2)}$ and $n^{(2)}$ as well.

In the present case, for every finite value of γ, the real and the imaginary part of e_i^+ are anti-correlated through $m^{(1)}$ (Equation 134). Correlations tend to align the network response on two main and opposite phases, as shown in the phase histograms of Figures S8C and S8D. The distribution of phases becomes sharper and sharper in the $\gamma \to 2$ limit, as the distribution in the complex plane collapses on the real axis.

The phase distribution across the population is reflected in the shape of the closed orbit defined by activity on the $m^{(1)} - m^{(2)}$ plane, whose components are given by κ_1 and κ_2. The phase of the oscillations in κ_1 (resp. κ_2) can be computed by projecting the eigenvector e^+ on the right-connectivity vectors $n^{(1)}$ and $n^{(2)}$:

$$\begin{aligned} \kappa_1 &= |\kappa_1|\, e^{i(\Phi_1 + \omega_H t)} + c.c. = \left\langle n_i^{(1)}[\phi_i] \right\rangle \\ \kappa_2 &= |\kappa_2|\, e^{i(\Phi_2 + \omega_H t)} + c.c. = \left\langle n_i^{(2)}[\phi_i] \right\rangle \end{aligned} \tag{136}$$

By using Equations 134 and 135 we get, in the linear regime:

$$\begin{aligned} \kappa_1 &= \left[\left\langle n_i^{(1)} m_i^{(2)} \right\rangle - \frac{\gamma}{2} \left\langle n_i^{(1)} m_i^{(1)} \right\rangle + i \left\langle n_i^{(1)} m_i^{(1)} \right\rangle \sqrt{\left| 1 - \frac{4}{\gamma^2} \right|} \right] e^{i\omega_H t} + c.c. \\ &= \left[\rho^2 \left(1 - \frac{\gamma^2}{2} \right) + i\gamma\rho^2 \sqrt{\left| 1 - \frac{4}{\gamma^2} \right|} \right] e^{i\omega_H t} + c.c. \end{aligned} \tag{137}$$

while:

$$\begin{aligned} \kappa_2 &= \left[\left\langle n_i^{(2)} m_i^{(2)} \right\rangle - \frac{\gamma}{2} \left\langle n_i^{(2)} m_i^{(1)} \right\rangle + i \left\langle n_i^{(2)} m_i^{(1)} \right\rangle \sqrt{\left| 1 - \frac{4}{\gamma^2} \right|} \right] e^{i\omega_H t} + c.c. \\ &= \left[\rho^2 \frac{\gamma}{2} - i\rho^2 \sqrt{\left| 1 - \frac{4}{\gamma^2} \right|} \right] e^{i\omega_H t} + c.c. \end{aligned} \tag{138}$$

When γ is close to 2, the complex amplitudes of κ_1 and κ_2 vanish. However, their real parts have different signs. We thus get: $\Phi_2 = 0$, $\Phi_1 = \pi$. As a consequence, at large γ values, the oscillatory activity in κ_1 and κ_2 tends to be strongly in anti-phase.

Stationary solutions can be instead easily analyzed with the standard mean-field approach. The equations for the first order statistics read:

$$\begin{aligned} \kappa^1 &= \left(\gamma\rho^2\kappa^1 + \rho^2\kappa^2 \right) \left\langle [\phi_i'] \right\rangle \\ \kappa^2 &= -\rho^2\kappa^1 \left\langle [\phi_i'] \right\rangle. \end{aligned} \tag{139}$$

The two equations can be combined together to give the following condition on $\langle[\phi_i']\rangle$, which in turn determines the value of Δ_0:

$$\rho^4 \left\langle [\phi_i'] \right\rangle^2 - \gamma\rho^2 \left\langle [\phi_i'] \right\rangle + 1 = 0. \tag{140}$$

The mean-field equations thus admit two solutions, given by:

$$\left\langle [\phi_i'] \right\rangle_\pm = \frac{\gamma}{2\rho^2} \left(1 + \pm \sqrt{1 - \frac{4}{\gamma^2}} \right) \tag{141}$$

which, similarly to Equation 133, take real values for $\gamma > 2$. Because of the constraints on the sigmoidal activation function, the mean-field solutions are acceptable only if $\left| \langle [\phi_i'] \rangle \right| < 1$. As it can be easily checked, the condition $\langle [\phi_i'] \rangle_- < 1$ coincides with imposing $\lambda_+ > 1$. We conclude that two stationary solutions exist above the instability boundary of the trivial fixed point (Figure S8A, gray). A second pair of solutions appears for $\langle [\phi_i'] \rangle_+ < 1$, which coincide with $\lambda_- > 1$ (Figure S8A, dashed), where the second outlier of J_{ij} becomes unstable. This second pair of solutions is however always dynamically unstable, as it can be checked by evaluating the outliers of

their stability matrix through Equation 113. The coefficients of the reduced matrix $\mathcal{M}$ read:

$$\begin{aligned} a_{11} &= \gamma\rho^2\langle[\phi_i']\rangle \\ a_{12} &= \rho^2\langle[\phi_i']\rangle \\ b_1 &= \frac{1}{2}\rho^2\left(\kappa^{20}+\gamma\kappa^{10}\right)\langle[\phi_i'']\rangle \end{aligned} \tag{142}$$

and

$$\begin{aligned} a_{21} &= -\rho^2\langle[\phi_i']\rangle \\ a_{22} &= 0 \\ b_2 &= -\frac{1}{2}\rho^2\kappa^{10}\langle[\phi_i'']\rangle. \end{aligned} \tag{143}$$

On the phase diagram boundary corresponding to $\gamma = 2$, the stable and the unstable pair of stationary solutions annihilate and disappear. At slightly smaller values of γ ($\gamma \lesssim 2$), the network develops highly non-linear and slow oscillations which can be thought of as smooth jumps between the two annihilation points (Figure S8D).

IMPLEMENTATION OF COMPUTATIONAL TASKS

Go-Nogo discrimination

Here we describe and analyze the unit-rank implementation of the Go-Nogo discrimination task (Figure 3).

The network receives inputs specified by N-dimensional vectors I^k. In every trial, the input vector coincides with one among the two vectors I^A and I^B, representing respectively the Go and the Nogo stimuli. The components of the two input patterns are generated independently from a Gaussian distribution of mean zero and variance Σ_I. As the components of the inputs are uncorrelated, the two vectors are mutually orthogonal in the limit of large N.

The network activity is readout linearly through a vector w generated from a Gaussian distribution of mean zero and variance Σ_w^2. The readout value is given by:

$$z = \frac{1}{N}\sum_{i=1}^{N} w_i\phi(x_i). \tag{144}$$

We fix the connectivity vectors m and n such that: (i) the readout is selective, i.e., $z \neq 0$ if the input is I^A and $z = 0$ for the input I^B; (ii) the readout is specific to the vector w, i.e., it is zero for any readout vector uncorrelated with w. The simplest network architecture which satisfies these requirements is given by:

$$\begin{aligned} m &= w \\ n &= I^A, \end{aligned} \tag{145}$$

i.e., the right-connectivity vector m corresponds to the readout vector, and the left-connectivity vector corresponds to the preferred stimulus I^A.

The response of the network can be analyzed by referring to the stationary and chaotic solutions of Equation 98. In the case analyzed here, the connectivity vectors have no overlap direction, so we set $M_m = M_n = M_I = \Sigma_{mI} = 0$, which implies $\mu = 0$. The first-order network statistics are determined by the overlap Σ_{nI} between the left-connectivity vector and the input vector. As the left-connectivity is given by I^A, Σ_{nI} is the overlap between the current input pattern I and the preferred pattern I^A, and it takes values $\Sigma_{nI} = \Sigma_I^2$ during the Go stimulus presentation and $\Sigma_{nI} = 0$ otherwise. From Equation 94 we have:

$$\begin{aligned} \kappa &= \langle n_i[\phi_i]\rangle \\ &= \langle I_i^A[\phi_i]\rangle. \end{aligned} \tag{146}$$

As a consequence, when the Go stimulus is presented ($I = I^A$):

$$\kappa = \Sigma_I^2\langle[\phi_i']\rangle, \tag{147}$$

while the first-order statistics κ vanishes in response to any orthogonal pattern I^B.

When activity is read out by the specific decoding vector w, the readout value is:

$$
\begin{aligned}
z &= \langle w_i[\phi_i] \rangle \\
&= \left\langle w_i \int \mathcal{D}z\phi\left(m_i\kappa + I_i + \sqrt{\Delta_0^I}z\right)\right\rangle \\
&= \left\langle w_i \int \mathcal{D}z\phi\left(w_i\kappa + I_i + \sqrt{\Delta_0^I}z\right)\right\rangle \\
&= \kappa\Sigma_w^2\left\langle\left[\phi_i'\right]\right\rangle,
\end{aligned}
\tag{148}
$$

while we trivially obtain $z = 0$ for any decoding set orthogonal to both connectivity vectors m and n.

In Figure 3C, we display the transient dynamics predicted by the mean-field theory within the $m - I$ plane. In order to compute the predicted trajectory, we use Equations 103 and 104, where the slowest time-scale of κ is computed by diagonalizing the reduced stability matrix in Equation 99.

In Figure 3G, we test the generalization properties of a network which responds to two Go patterns I_1^A and I_2^A. We examine the response to a normalized mixture input defined as:

$$
I = \sqrt{\alpha}I_1^A + \sqrt{1-\alpha}I_2^A, \tag{149}
$$

so that the variance of the total input is fixed and equal to Σ_I^2. We set $n = I_1^A + I_2^A$, so that the equation for the first-order statistics reads:

$$
\begin{aligned}
\kappa &= \left\langle I_{1i}^A[\phi_i]\right\rangle + \left\langle I_{2i}^A[\phi_i]\right\rangle \\
&= \left(\sqrt{\alpha} + \sqrt{1-\alpha}\right)\Sigma_I^2\left\langle\left[\phi_i'\right]\right\rangle.
\end{aligned}
\tag{150}
$$

Detection of a continuous noisy stimulus

In Figure 4, we construct a network model which performs a Go-Nogo detection task on a one-dimensional continuous stimulus.

The stimulus consists of an input of time-varying amplitude $c(t)I$. As in Figure 3, the input direction I is a centered Gaussian vector of variance Σ_I^2. The strength value $c(t)$ includes a stationary component $\overline{c}$ together with additive white noise of standard deviation σ. Less importantly, we include in the input an orthogonal component of quenched noise of unitary variance. The network output is defined at the level of an orthogonal readout as in Equation 144, and the task consists in responding to the stimulus when the strength of the input c is larger than a given threshold.

We obtain highly non-linear readout responses by considering non-vanishing overlaps between the connectivity vectors m and n. The simplest setup consists of taking:

$$
\begin{aligned}
m &= w + \rho_m y \\
n &= I + \rho_n y,
\end{aligned}
\tag{151}
$$

where y is a standard Gaussian vector which defines a direction common to m and n, but orthogonal both to w and I.

For this configuration, as in Equation 94, the mean-field equation for the first-order statistics includes two terms, generated respectively by the input and the rank-one structure:

$$
\kappa = \left(\rho_m\rho_n\kappa + \overline{c}\,\Sigma_I^2\right)\left\langle\left[\phi_i'\right]\right\rangle. \tag{152}
$$

Before the stimulus presentation ($\overline{c} = 0$, $\sigma = 0$), the structure overlap $\rho_m\rho_n$ is strong enough to generate two bistable solutions (Figure 1). We set the negative κ solution to represent the Nogo condition, and we initialize the network in this state. To have a zero output in this condition, we add an offset to the readout.

When an input along the preferred direction is presented ($\overline{c} > 0$), two asymmetric solutions exist only when the strength of the input $\overline{c}$ is not too large (Figure 2D, right). When the correlation $\overline{c}$ is large, instead, only the positive branch of the solution is retrieved (Figure 2D, right). As a consequence, the average value of κ (and thus the readout signal) jumps to positive values, which define the Go output condition.

More generally, in order to compute the network performance (Figure 4G), the network is said to respond to the stimulus if the readout z at the end of the stimulus presentation takes values larger than one half of the readout value expected for the upper state.

The threshold value for $\overline{c}$ at which the bistability disappears is mostly determined by the strength of the structure overlap, but depends also the input and readout parameters Σ_I and Σ_w. For practical purposes, in order to obtain the model implementation illustrated in Figure 4, we first fix the values of $\Sigma_I = 1.2$, $\Sigma_w = 1.2$ and $\rho_n = 2$. We then tune the value of ρ_m in order to obtain a threshold value for c close to 0.5. This leads to $\rho_m = 2$.

In Figure 4F we vary ρ_m and we show that the value of the threshold decreases to zero as the structure strength $\rho_m\rho_n$ decreases from its original value ($\rho_m\rho_n = 4$). Rank-one structures characterized by different strengths thus correspond to different thresholds, but also induce different dynamical time-scales in the network. As a rough estimate of this time-scale, we compute the inverse of the outlier eigenvalue from the stability matrix of the fixed point corresponding to the Go resting state ($\overline{c} = 0$). The value of the outlier can be computed from the linearized mean-field equations (Equation 71). We show that arbitrarily large time-scales are only obtained

by decreasing the value of the structure strength to the critical point where the two bistable branches of the solution emerge from the trivial fixed point. In this configuration, the threshold detected by the network is arbitrarily small.

Contextual modulation of threshold value

Here we briefly illustrate how the threshold of detection can be controlled by an additional modulatory input (Figure 5B). Modulatory inputs are used in Figures 5 and 6 to implement more complex tasks which require context-dependent responses to stimuli. Any input direction which overlaps with the left-connectivity vector n and is orthogonal to the stimulus axis I can serve as modulatory input. For simplicity, we consider modulatory inputs which are aligned with the overlap direction y (see Equation 151). The total external input to the network contains the modulatory component γy together with the stimulus term $c(t)I$, where γ is a scalar which controls the strength of the modulation. The mean-field equation for the first-order statistics reads:

$$\kappa = \left(\rho_m \rho_n \kappa + \rho_n \gamma + \overline{c}\ \Sigma_I^2\right) \left\langle \left[\phi_i'\right] \right\rangle. \tag{153}$$

Equation 153 indicates that the modulatory component of the input acts as a constant offset to the stimulus strength. Its net effect is to shift the response curve of the network along the x axis (Figure 5B) by an amount directly regulated by the parameter γ. Varying γ thus results in network models which detect variable threshold values.

Rank-two structures for context-dependent computations

Here we provide details on the rank-two implementation of the context-dependent tasks. The same model has been used for both tasks in Figures 5 and 6.

The stimuli consist of combinations of two different features A and B that correspond to inputs along two directions I^A and I^B, generated as Gaussian random vectors of variance Σ_I^2. Contextual cues are represented as additional inputs along directions I_{ctxA} and I_{ctxB} of unit variance. The total input pattern to the network on a given trial is therefore given by:

$$I(t) = c_A(t) I^A + c_B(t) I^B + \gamma_A I_{ctxA} + \gamma_B I_{ctxB}. \tag{154}$$

The values c_A and c_B express the strength of the stimulus along the two feature directions. They are given by the sum of stationary average values ($\overline{c}_A, \overline{c}_B$), and temporary fluctuating components generated from independent realizations of white noise with standard deviation σ. In the simple discrimination version of the task (Figure 5), inputs are noise-free ($\sigma = 0$) and consist of a single feature in each trial ($\overline{c}_A = 1$ and $\overline{c}_B = 0$ or vice versa). In the evidence integration version of the task (Figure 6), inputs are noisy ($\sigma > 0$) and include non-zero average components along both feature directions. Finally, the parameters γ_A and γ_B control the two modulatory inputs which are taken in the directions defined by I_{ctxA} and I_{ctxB}.

In order to implement context-dependent computations, we define a unique readout signal $z(t)$ by using a common readout set w of unit variance (Equation 144), to which we add an offset so that the baseline Nogo output is set to zero. The network is said to respond to the stimulus if the value of the total readout at the end of the stimulus presentation takes values larger than one half of the largest predicted value for the upper state.

The rank-two connectivity matrix we consider is given by:

$$\begin{aligned} m^{(1)} &= y_A + \rho_m I_{ctxA} + \beta_m w \\ n^{(1)} &= I^A + \rho_n I_{ctxA} + \beta_n w \\ m^{(2)} &= y_B + \rho_m I_{ctxB} + \beta_m w \\ n^{(2)} &= I^B + \rho_n I_{ctxB} + \beta_n w, \end{aligned} \tag{155}$$

where vectors y_A and y_B represent the orthogonal components of the right-connectivity vectors and are generated as Gaussian vectors of fixed variance (for simplicity, we set $\Sigma_y = \Sigma_I$).

For our choice of the parameters, the network solves the two different tasks by relying on the strongly non-linear responses generated by the interplay between the recurrent connectivity and the feed-forward inputs (details given below).

For weak input values, the network dynamics is characterized by two stable attractors (Figure 6F). As in Figure 4, we initialize the network in the state characterized by negative κ_1 and κ_2 values before the stimulus presentation. This dynamical attractor corresponds to the Nogo state. For strong input strengths, the network can jump to the Go state, defined as the stable attractor characterized by positive κ_1 and κ_2 values.

The rank-two connectivity matrix has been designed as an extension of the unit-rank recurrent connectivity employed in Figure 4. We started by setting:

$$\begin{aligned} m^{(1)} &= y_A + \rho_m I_{ctxA} \\ n^{(1)} &= I^A + \rho_n I_{ctxA} \\ m^{(2)} &= y_B + \rho_m I_{ctxB} \\ n^{(2)} &= I^B + \rho_n I_{ctxB}. \end{aligned} \tag{156}$$

Note that, because the only overlap directions (I_{ctxA} and I_{ctxB}) are internal to the $m^{(1)} - n^{(1)}$ and $m^{(1)} - n^{(1)}$ pairs, Equation 156 describes a rank-two structure which generates a continuous ring attractor as in Figures S5D–S5I (gray circles in Figure 6F).

The readout $z(t)$ should detect the presence of both stimuli directions. As a consequence, it should be sensitive to both overlap values κ_1 and κ_2. For this reason, we introduce a common term in the four connectivity vectors that is aligned to the common readout (Equation 155).

Introducing a common overlap direction has the effect of destabilizing the continuous attractor dynamics along the direction $\kappa_1 = \kappa_2$ (dashed line in Figure 6F), where two stable and symmetric fixed points are generated. The equations for the first-order spontaneous dynamics read indeed:

$$\begin{aligned}\kappa_1 &= \langle n^{(1)}[\phi_i]\rangle = \rho_m\rho_n\kappa_1\langle[\phi_i']\rangle + \beta_m\beta_n(\kappa_1+\kappa_2)\langle[\phi_i']\rangle \\ \kappa_2 &= \langle n^{(2)}[\phi_i]\rangle = \rho_m\rho_n\kappa_2\langle[\phi_i']\rangle + \beta_m\beta_n(\kappa_1+\kappa_2)\langle[\phi_i']\rangle\end{aligned} \tag{157}$$

from which the value of $\kappa_1 = \kappa_2 = \bar{\kappa}$ can be derived by dividing and multiplying together the two equations. The final readout signal contains a contribution from both first-order statistics:

$$z(t) = \langle w_i[\phi_i]\rangle = \beta_m(\kappa_1+\kappa_2)\langle[\phi_i']\rangle. \tag{158}$$

The input-driven dynamics of the network are determined by the interplay between the structure strength and the contextual and stimulus inputs. Crucially, the modulatory inputs along I_{ctxA} and I_{ctxB} are used to gate a context-dependent response. Similarly to Figure 5B, a strong and negative gating variable along I_{ctxA} can completely suppress the response to stimulus I^A, so that the readout signal is left free to respond to I^B.

The overall effects of the inputs on the dynamics can be quantified by solving the mean-field equations. For the first-order statistics, we obtain:

$$\begin{aligned}\kappa_1 &= \langle[\phi_i']\rangle\{\rho_m\rho_n\kappa_1 + \beta_m\beta_n(\kappa_1+\kappa_2) + \bar{c}_A\Sigma_I^2 + \rho_n\gamma_A\} \\ \kappa_2 &= \langle[\phi_i']\rangle\{\rho_m\rho_n\kappa_2 + \beta_m\beta_n(\kappa_1+\kappa_2) + \bar{c}_B\Sigma_I^2 + \rho_n\gamma_B\}\end{aligned} \tag{159}$$

while the second-order gives, in the case of stationary regimes:

$$\Delta_0 = g^2\langle[\phi_i^2]\rangle + \Sigma_w^2(\kappa_1^2+\kappa_2^2) + \beta_m^2(\kappa_1^2+\kappa_2^2) + \Sigma_I^2(\bar{c}_A^2+\bar{c}_B^2) + (\rho_m\kappa_1+\gamma_A)^2 + (\rho_m\kappa_2+\gamma_B)^2. \tag{160}$$

Figures S5L–S5M displays the values of the first-order statistics and the readout response in the two contexts. Note that, when the response to I^A (resp. I^B) is blocked at the level of the readout, the relative first-order statistics κ_1 (resp. κ_2) does not vanish, but actively contributes to the final network response.

The average activation variable of single neurons contains entangled contributions from the main directions of the dynamics, which are inherited both from the external inputs and the recurrent architecture:

$$\mu_i = [x_i] = (y_{A,i} + \rho_m I_{ctxA,i} + \beta_m w_i)\kappa_1 + (y_{B,i} + \rho_m I_{ctxB,i} + \beta_m w_i)\kappa_2 + \bar{c}_A I_i^A + \bar{c}_B I_i^B + \gamma_1 I_{ctxA,i} + \gamma_2 I_{ctxB,i}. \tag{161}$$

In Figures 5E and 6D, we project the averaged activation μ_i in the directions that are more salient to the task. The projection along w, which reflects the output decision, is proportional to the readout value (Equation 158). The input signals affect instead the average activity through the values of κ_1 and κ_2, but can be also read out directly along the input directions. Note that the projection on the input direction I^A (resp. I^B) is proportional to the signal $\bar{c}_A$ (resp. $\bar{c}_B$) regardless of the configuration of the modulatory inputs selecting one input channel or the other.

In practical terms, in order to obtain the network architecture that has been used in Figures 5 and 6, we fixed the parameters step by step. We first considered input patterns only along I^A ($\bar{c}_B = 0$), and we fixed two arbitrary values of β_m and β_n. In particular, we considered intermediate values of β. Large values of β tend to return large activity variance, which requires evaluating with very high precision the Gaussian integrals present in the mean-field equations. Small values of β bring instead the network activity closer to a continuous-attractor structure, and turn into larger finite-size effects. In a second step, we fix ρ_m and ρ_n such that the network detects normalized input components along I^A only when they are larger than a threshold value, that is taken around 0.5. We then looked for a pair of gating variables strengths $[\gamma_A, \gamma_B]$ which completely suppresses the response to I^A by extending the range of bistable activity. The opposite pattern can be used to block the response in I^B and allow a response in I^A.

Once the response in I^A has been blocked, it can be verified that the network solely responds to inputs which contain a response along I^B that is larger than a threshold close to 0.5. Note that, as in Figures S5L–S5M, different values of $\bar{c}_A$ only minimally affect the exact position of the threshold.

To conclude, we remark that this procedure leaves the freedom of fixing the network parameters in many different configurations. The complex rank-two architecture leads to larger finite-size effects than the respective unit-rank setup which acts as a single detector of correlations. In particular, the error at the level of the readout is larger but it decays with the system size, as expected for deviations induced by finite-size effects (Figure S5N). Finally, note that when the noise in the input stimuli becomes extremely large, the network loses its ability to respond in a totally context-dependent fashion, as strong fluctuations in the non-relevant stimulus become likely to elicit a response.

METHOD DETAILS FOR MAIN FIGURES

Figure 1

In this figure, $\Sigma_m = \Sigma_n = 1.0$. Note that the precise position of the instability to chaos depends on the value of Σ_m. The connectivity vectors m and n were generated from bivariate Gaussian distributions (means M_m and M_n, variances Σ_m and Σ_n, correlation ρ). Here we display the case where m and n overlap only along the unitary direction ($M_m > 0$, $M_n > 0$, $\rho = 0$, see STAR Methods). As shown in Figure S2, qualitatively similar regimes are obtained when the overlap is defined on an arbitrary direction. C-D: Network simulations were performed starting from initial conditions centered around m and $-m$. Activity is integrated up to $T = 800$. In simulations, $N = 5000$, and statistics are averaged over 15 different connectivity realizations. The error bars, when visible, correspond to the standard deviation of the mean (as in every other figure, if not differently specified).

Figure 2

In this figure, $g = 0.8$. Other parameters are set as in Figure 1. B: The asymptotic input parameters are indicated by gray dots in D (middle). The simulation results (dark gray traces) correspond to 20 trajectories for different network realizations (different trajectories strongly overlap). We simulated $N_{tr} = 20$ different networks, each consisting of $N = 3500$ units. In every network realization, the random part of the connectivity χ_{ij} is varied, while the low-rank part $m_i n_j$ is kept fixed. I (resp. m) scale: 0.7 (resp. 0.25) units. D: The external input is increased along $n_\perp$, the component of n that is perpendicular to the overlap direction.

Figure 3

The input and the readout vectors are Gaussian patterns of standard deviation $\Sigma = 2$. C (right): Colored traces: 20 trajectories from different network realizations (different trajectories strongly overlap). We simulated $N_{tr} = 20$ different realizations of the network, each consisting of $N = 2500$ units. In every network realization, the random part of the connectivity χ_{ij} is generated independently, while the low-rank part $m_i n_j$ is kept fixed. I^A, I^B and m scale: 1.5 units. D: Here, and in every plot if not differently stated, ρ indicates the Pearson correlation coefficient. F: The PC axis are determined by analyzing separately the trials corresponding to the Go (top) and the Nogo (bottom) stimuli. Connectivity is measured as the average reciprocal synaptic strength; it includes both the random and the unit-rank components and it is averaged across network realizations. Note that the value of the correlation coefficient ρ increases with the number of realizations N_{tr} and the structure strength.

Figure 4

The input and the readout vectors are Gaussian patterns of standard deviation $\Sigma = 1.2$. The overlap between the connectivity vectors m and n leading to non-linear responses is quantified by $\rho_m = \rho_n = 2.0$. B: The input is generated as white noise of mean $\overline{c} = 0.6$ and standard deviation $\sigma = 0.4$ (the noise trace in the figure is only for illustration purposes). The red dashed line indicates the threshold in the implemented network. C: The gray bar indicates the time point at which the network output is measured. Here and in the following figures, the readout includes an offset, so that the baseline value is set to zero. D: We simulated many input noise traces for $N_{tr} = 4$ different realizations of the network, each consisting of $N = 2500$ units. In every network realization, the random part of the connectivity χ_{ij} is varied, while the low-rank part $m_i n_j$ is kept fixed. Trajectories are smoothed with a Gaussian filter of standard deviation equal to one normalized time unit. I (resp. m) scale: 0.5 (resp. 3.5) units. F: The structure strength corresponds to the overlap $\rho_m \rho_n$. The effective timescale is measured as the inverse of the value of the outlier eigenvalue of the stability matrix for $\overline{c} = 0$. G: The psychometric curve was measured across $N_{tr} = 100$ different realizations. The network produces an output to the stimulus if at the end of the stimulus presentation (vertical gray line in B) the value of the readout z is larger than one half of the largest readout value predicted by the theory. H: Details as in Figure 3F.

Figure 5

The stimuli vectors are Gaussian patterns of standard deviation $\Sigma = 1.2$. We furthermore set: $g = 0.8$, $\beta_m = 0.6$, $\beta_n = 1$, $\rho_m = 3$, $\rho_n = 1.6$. The amplitudes of the two context directions are fixed to $[0.08, -0.14]$ (resp. $[-0.14, 0.08]$) during the context A (resp. context B) trials. B: We consider in this case a unit-rank network as in Figure 2D, and we show in the two panels the network response for two different values of the input strength along the overlap axis (we set, respectively, $M_I = -0.3$ and 0.6). Details on the effect of contextual modulation on the full rank-two model are further illustrated in Figures S5L–S5N. E: We simulated $N_{tr} = 4$ different realizations of the network, each consisting of $N = 3000$ units. In every network realization, the random part of the connectivity χ_{ij} is varied, while the low-rank part $m_i n_j$ is kept fixed. I^A and I^B (resp. w) scale: 1.0 (resp. 2.0) units. F: The network performance was measured across $N_{tr} = 50$ different network realizations of size $N = 7500$. The network produces an output to the stimulus if at the end of the stimulus presentation (vertical gray line in D) the value of the readout z is larger than one half of the largest readout value predicted by the theory. G: Details as in Figure 3F.

Figure 6

The stimuli vectors are Gaussian patterns of standard deviation $\Sigma = 1.2$. We furthermore set: $g = 0.8$, $\beta_m = 0.6$, $\beta_n = 1$, $\rho_m = 3$, $\rho_n = 1.38$. The amplitudes of the two context directions are fixed to $[0.08, -0.18]$ (resp. $[-0.18, 0.08]$) during the context A (resp. context B) trials. B: Here $\overline{c}_A = 0.6$ and $\overline{c}_B = 0.1$, while the standard deviation of the noise in the input is $\sigma = 0.3$ (the noise trace in

the figure is only for illustration purposes). D: We simulated many noisy input traces for $N_{tr} = 5$ different realizations of the network, each consisting of $N = 4000$ units. In every network realization, the random part of the connectivity χ_{ij} is varied, while the low-rank part $m_i n_j$ is kept fixed. For the sake of clarity, only correct trials have been included. I^A and I^B (resp. w) scale: 1 (resp. 1.5) units. E: Network performance was measured across $N_{tr} = 50$ different network realizations of size $N = 7500$.

QUANTIFICATION AND STATISTICAL ANALYSIS

In this section, we briefly describe the analysis techniques that have been applied to the datasets generated from direct simulations of activity in finite-size networks (Figures 2, 3, 4, 5, and 6).

Dimensionality reduction

In order to extract from the high-dimensional population activity the low-dimensional subspace which contains most of the relevant dynamics, we performed dimensionality reduction via a standard Principal Component (PC) analysis.

To begin with, we constructed the activation matrix X. In X, every column corresponds to the time trace of the activation variable $x_i(t)$ for unit i, averaged across trials. We indicate as *trials* different network simulations, where different noisy inputs, or different quenched noise in the random connectivity matrix have been generated (details are specified in the figure captions). The activation matrix X is normalized through Z-scoring: to every column, we subtract its average over time, and we divide by its standard deviation. Note that Z-scoring distorts the shape of the population trajectory in the phase space. For this reason, in order to facilitate the comparison with the trajectory predicted by the mean-field theory, in Figure S3 we more simply consider the mean-subtracted matrix X. Applying the PCA analysis to one of the two data formats impacts the results from a quantitative point of view, but does not change their general validity.

The principal components (PC) are computed as the normalized eigenvectors $\{e_l\}_{l=1,\ldots,N}$ of the correlation matrix $C = X^T X$. The PC are sorted in decreasing order according to the corresponding real eigenvalue λ_l. The activation matrix X can be projected on the orthonormal basis generated by the PC vectors by computing: $X' = XE$, where E is the $N \times N$ matrix containing the PC eigenvectors ordered as columns. The variance explained by the l-th PC mode e_l can be computed as the l-th entry on the diagonal of the rotated correlation matrix $C' = X'^T X'$.

While in our network models the low-rank part of the connectivity determines a purely low-dimensional dynamics (Figure S3A), the random part of the connectivity generates a continuum of components whose amplitude is determined by strength of the random connectivity g with respect to the connectivity and input vectors. In Figure 2, where $g = 0.8$, the low-dimensional nature of the dynamics is revealed by considering averages across several $(N_{tr} = 20)$ realizations of the random connectivity. In Figure S3B, we illustrate the result of performing PCA on the activity generated by a single network. In this case, even if more PC components contribute to the total variance, the two first axis bear a strong resemblance with the directions predicted with the theory. In Figure S3C we show that, in the same spirit, a PCA analysis can be used to extract the relevant geometry of the network model also when activity is strongly chaotic.

In order to more easily connect with the theoretical predictions, we systematically applied dimensionality reduction on datasets constructed from the activation variable x_i. We verified that our results still hold, from a qualitative point of view, when the analysis is performed on the non-linearly transformed variables $\phi(x_i)$. In the network models we considered, the activation variables $\phi(x_i)$ indeed form a non-linear but dominantly low-dimensional manifold in the phase space. The axes predicted by the mean-field theory determine the dominant linear geometry of this manifold, and can be still captured (although less precisely) by looking at the first PC components.

Linear regression

In order to estimate how single units in the network are tuned to different task variables (such as input stimuli or decision variables), we used a multi-variate linear regression analysis.

To this end, we considered the full population response $x_i^k(t)$, where $k = 1, \ldots, N_{tr}$ indicates the trial number. Following (Mante et al., 2013), our aim was to describe the network activation variables as linear combinations of the M relevant task variables. In Figure 3, the two variables we considered were the strength of the Go and of the Nogo inputs, that we indicate here with c_{Go} and c_{Nogo}:

$$x_i^k(t) = \beta_{i,t}^{\text{Go}} c_{\text{Go}}(k) + \beta_{i,t}^{\text{Nogo}} c_{\text{Nogo}}(k). \tag{162}$$

In a Go, or in a Nogo trial, only one of the two strength coefficients is non-zero. In Figure 4, the two relevant task variables are assumed to be the input strength along I, quantified by c, and the network output, quantified as the value of the readout z at the end of the stimulus presentation:

$$x_i^k(t) = \beta_{i,t}^{\text{input}} c(k) + \beta_{i,t}^{\text{choice}} z(k). \tag{163}$$

In Figures 5 and 6, the relevant variables are four: the strength of stimuli *A* and *B*, the trial context and the network output. We thus have:

$$x_i^k(t) = \beta_{i,t}^A c_A(k) + \beta_{i,t}^B c_B(k) + + \beta_{i,t}^{\mathrm{ctx}} y(k) + \beta_{i,t}^{\mathrm{choice}} z(k). \tag{164}$$

where the context variable is represented by a unique symbolic variable *y*, which takes value $y = 1$ in context *A* and $y = -1$ in context *B*.

More generally, we indicate with $\beta_{i,t}^{\nu}$ the regression coefficient of unit *i* with respect to the task feature ν at time *t*. The vector $\beta_{i,t} = \{\beta_{i,t}^{\nu}\}_{\nu=1,..,M}$ indicates the collection of the *M* variables regressors for a given unit at the time point *t*. We compute the regression coefficients by defining a matrix *F* of size $M \times N_{tr}$, where every row contains the value of the *M* relevant task variables across trials. The regression coefficient vectors are then estimated by least-square inversion:

$$\beta_{i,t} = \left(FF^T\right)^{-1} F x_{i,t} \tag{165}$$

where the vector $x_{i,t}$ is constructed by collecting across trials the value the activation variable of unit *i* at time *t*.

In order to get rid of the time dependence of our result, we simply consider the coefficients $\beta_{i,t}$ at the time point where the two-dimensional array $\beta_{i,t}$ for every *i* has maximal norm (Mante et al., 2013). The resulting set of *M*-dimensional vectors β_i contains the regression coefficients of unit *i* with respect to the *M* relevant task variables. The *N*-dimensional regression axis for a given task variable ν is finally constructed by collecting the ν-th components of β_i across different population units: $\{\beta_i^{\nu}\}_{i=1,..,N}$.

DATA AND SOFTWARE AVAILABILITY

Software was written in the Python (http://python.org) programming languages. Implementations of algorithms used to compute quantities presented in this study are available at: https://github.com/fmastrogiuseppe/lowrank/.