

On the relationship between equilibria and dynamics in large, random neuronal networks

Xiaoyu Yang,^{1,2,3} Giancarlo La Camera,^{2,3,4,5,6,*} and Gianluigi Mongillo^{7,8,9,†}

¹Graduate Program in Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA

²Department of Neurobiology and Behavior, Stony Brook University, Stony Brook, NY, USA

³Center for Neural Circuit Dynamics, Stony Brook University, Stony Brook, NY, USA

⁴AI Innovation Institute, Stony Brook University, Stony Brook, NY, USA

⁵Center for Advanced Computational Science, Stony Brook University, Stony Brook, NY, USA

⁶Graduate Programs in Neuroscience, Stony Brook University, Stony Brook, NY, USA

⁷Sorbonne Université, INSERM, CNRS, Institut de la Vision, F-75012 Paris, France

⁸Centre National de la Recherche Scientifique, Paris, France

⁹School of Natural Sciences, Institute for Advanced Study, Princeton, NJ, USA.

We investigate the equilibria of a random model network exhibiting extensive chaos. In this regime, a large number of equilibria is present. They are all saddles with low-dimensional unstable manifolds. Surprisingly, despite network's connectivity being completely random, the equilibria are strongly correlated and, as a result, they occupy a very small region in the phase space. The attractor is inside this region. This geometry explains why the collective states sampled by the dynamics are dominated by correlation effects and, hence, why the chaotic dynamics in these models can be described by a fractionally-small number of collective modes.

Large systems of non-linear differential equations are a standard modeling framework for many natural and technological systems. Qualitative analysis, however, is of limited help in the study of their dynamics, because of the large number of degrees of freedom. One general approach, for suitably idealized systems, is dynamical mean-field theory (DMFT) [1, 2]. DMFT allows one to obtain a drastic, and exact, dimensionality reduction; one only needs to consider a single degree of freedom interacting with an effective ‘field’, to be determined self-consistently. Unfortunately, though exact, the DMFT equations are usually still analytically intractable.

These difficulties have motivated the search for approaches complementary to DMFT. In particular, a promising approach focuses on determining the number of equilibria and their stability [3–7]. The hope is to use this information to infer qualitative/quantitative aspects of the dynamics, without solving the DMFT [3, 6, 7]. Alternatively, this information could conceivably help in solving the DMFT numerically.

Previous studies using this approach did not investigate the spatial organization of the equilibria, and only one [6] attempted to *locate* the dynamics with respect to the equilibria. In this letter, we address these issues by considering a random neuronal network exhibiting the paradigmatic transition between a unique, stable equilibrium and a chaotic attractor [8]. Our model network is as follows. The instantaneous state of neuron i ($i = 1, \dots, N$) is described by x_i , which evolves according to

$$\dot{x}_i = -x_i + \sqrt{N} + \frac{1}{\sqrt{N}} \sum_{j=1}^N w_{ij} \phi_j \equiv v_i(\mathbf{x}), \quad (1)$$

with $\phi_j \equiv \phi(x_j)$ and $\phi(x) \equiv x\Theta(x)$, where $\Theta(x)$ is the

Heaviside function. The term $\sqrt{N}$ is an external, excitatory input to the network. The w_{ij} 's are the synaptic efficacies, that we take as i.i.d. Gaussian deviates with $\mathbb{E}[w_{ij}] = -1$ and $\text{Var}[w_{ij}] = \sigma_w^2$. As we show below, in the limit $N \rightarrow \infty$, our model network exhibits a unique stable equilibrium for $\sigma_w < \sqrt{2}$ and chaotic dynamics for $\sigma_w > \sqrt{2}$ [9].

The number of network's hyperbolic equilibria for a given sample $\mathbf{w}$, $P_{\mathbf{w}}$, is given by the so-called Kac-Rice formula [10], i.e.,

$$P_{\mathbf{w}} = \int_{-\infty}^{+\infty} d\mathbf{x} |\det \mathbf{J}_{\mathbf{w}}(\mathbf{x})| \delta(\mathbf{v}_{\mathbf{w}}(\mathbf{x})), \quad (2)$$

where $\mathbf{J}_{\mathbf{w}}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{v}_{\mathbf{w}}(\mathbf{x})$, evaluated at $\mathbf{x}$. The subscript denotes dependence on the sample $\mathbf{w}$. $P_{\mathbf{w}}$ fluctuates from sample to sample even for $N \rightarrow \infty$. Roughly speaking, this is because $P_{\mathbf{w}}$ is a product, rather than a sum, of random variables. Accordingly, we expect $P_{\mathbf{w}} \sim \exp(N\Sigma_q^*)$ for N large, where

$$\Sigma_q^* = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log P_{\mathbf{w}}]. \quad (3)$$

The expectation is over $\mathbf{w}$. Thus, the *quenched* complexity, Σ_q^* , describes the asymptotic behavior of the typical value of $P_{\mathbf{w}}$. The *annealed* complexity, Σ_a^* , instead describes the asymptotic behavior of $\mathbb{E}[P_{\mathbf{w}}]$, i.e., $\mathbb{E}[P_{\mathbf{w}}] \sim \exp(N\Sigma_a^*)$. It is always $\Sigma_a^* \geq \Sigma_q^*$.

We evaluate Σ_q^* using the replica method [11]. The method relies on the computation of $\mathbb{E}[P_{\mathbf{w}}^n]$ for integer n , from which $\mathbb{E}[\log P_{\mathbf{w}}] = \lim_{n \rightarrow 0} n^{-1} \log \mathbb{E}[P_{\mathbf{w}}^n]$ (for n real). The major technical difficulty in computing $\mathbb{E}[P_{\mathbf{w}}^n]$ is to average the replicated determinant of $\mathbf{J}_{\mathbf{w}}(\mathbf{x})$ [4, 5].

That is because $|\det \mathbf{J}_{\mathbf{w}}(\mathbf{x})|$ is a highly non-linear function of the w_{ij} 's.

We note that, when $\mathbf{x}$ is independent of $\mathbf{w}$, the distribution of the eigenvalues of $\mathbf{J}_{\mathbf{w}}(\mathbf{x})$ is self-averaging [12]. It depends only on σ_w and $\frac{1}{N} \sum_i (\phi'_i)^2$, where ϕ'_i is short-hand for the derivative of $\phi(x)$ evaluated at x_i . In our case, $\phi'_i = \Theta(x_i)$. Hence, $\frac{1}{N} \sum_i (\phi'_i)^2 \equiv f$ is simply the fraction of ‘active’ neurons (i.e., $\phi_i > 0$) at $\mathbf{x}$. Then, $\mathbf{J}_{\mathbf{w}}(\mathbf{x})$ has one real eigenvalue $\simeq -\sqrt{N}$, $(1-f)N$ eigenvalues equal to -1 , and the density of the remaining fN eigenvalues, which henceforth we denote λ_k with $k = 1, \dots, fN$, is the uniform distribution within the disc of radius $R = \sigma_w \sqrt{f}$ centered at $-1 + i0$, which henceforth we denote $\rho(z; f, \sigma_w)$.

The equilibria $\mathbf{x}$ at which one needs to evaluate $|\det \mathbf{J}_{\mathbf{w}}(\mathbf{x})|$ in Eq. (2), however, *depend* on $\mathbf{w}$. To make progress, we assume that this dependence is sufficiently weak so that the density of the λ_k 's is given by $\rho(z; f, \sigma_w)$ also when $\mathbf{x}$ is an equilibrium of Eq. (1) [12]. Then, to leading order,

$$|\det \mathbf{J}_{\mathbf{w}}(\mathbf{x})| \sim \exp[N\zeta(f, \sigma_w)], \quad (4)$$

where $\zeta(f, \sigma_w) \equiv \int dz \rho(z; f, \sigma_w) \log|z|$. Using Eq. (4) in Eq. (2), we obtain

$$P_{\mathbf{w}} \sim \int_0^1 df \exp[N\zeta(f, \sigma_w)] \times \int_{-\infty}^{+\infty} d\mathbf{x} \delta\left(\sum_{i=1}^N \Theta(x_i) - fN\right) \delta(\mathbf{v}_{\mathbf{w}}(\mathbf{x})), \quad (5)$$

where the additional delta function enforces the constraint that the equilibria have a fraction f of active neurons. The only term that depends on $\mathbf{w}$ in Eq. (5) now is $\mathbf{v}_{\mathbf{w}}(\mathbf{x})$, and this dependence is linear in the w_{ij} 's.

Using Eq. (5), $\mathbb{E}[P_{\mathbf{w}}^n]$ can be evaluated by standard techniques [11]. This leads to the introduction of some auxiliary functions of f [11], in particular the *overlaps*, $q^{ab}(a, b = 1, \dots, n)$, defined as

$$q^{ab}(f) \equiv \lim_{N \rightarrow \infty} \frac{1}{N} (\phi^a \cdot \phi^b) = \sqrt{q^{aa}(f) q^{bb}(f)} \cos \theta^{ab}(f) \quad (6)$$

where $\phi^a \equiv (\phi(x_i^a))$, $(\cdot)$ denotes the scalar product, and $\mathbf{x}^a$ and $\mathbf{x}^b$ are two equilibria with the *same* f . The equality in Eq. (6) follows from $\|\phi^a\| \equiv \sqrt{\phi^a \cdot \phi^a} = \sqrt{N q^{aa}(f)}$ ($N \rightarrow \infty$); hence, $\cos \theta^{ab}(f)$ is the cosine similarity (CS) between ϕ^a and ϕ^b .

The q^{ab} 's, and the other auxiliary functions, are determined self-consistently by making an explicit assumption about their symmetry under permutations of the indexes ab (replica symmetry) [13]. We make the simplest assumption, the replica symmetric ansatz, for *all* values of

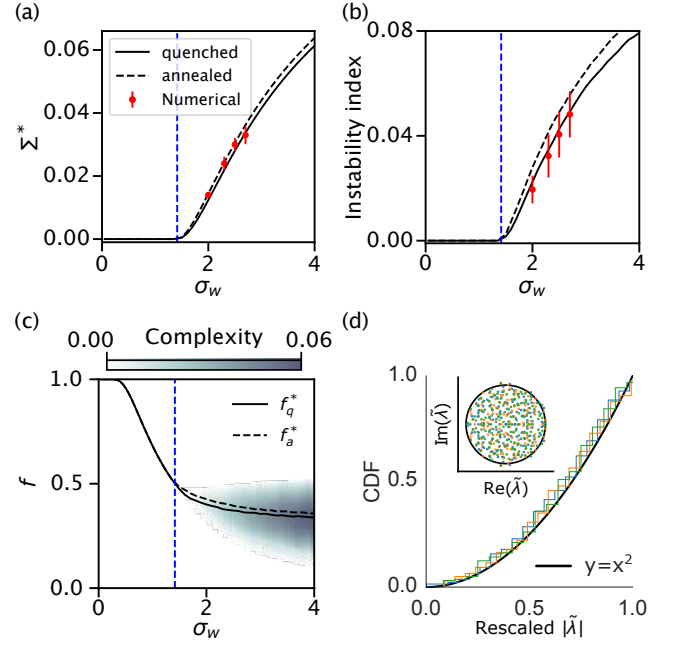


FIG. 1. (a) Σ_q^* (full) and Σ_a^* (dashed) as a function of σ_w . Dots with errorbars (SD) represent numerical estimates ($n = 50$ sample networks for each value of σ_w and N). The vertical line (dashed blue) indicates the critical point $\sigma_w = \sqrt{2}$. (b) Instability index averaged over all the equilibria as a function of σ_w , according to the quenched (full) and annealed (dashed) theory. Dots and vertical line as in (a). (c) Σ_q as a function of σ_w and f . The full line denotes the location of Σ_q^* , the dashed one that of Σ_a^* . Vertical dashed line as in (a). (d) Spectral distribution of $\tilde{\lambda} \equiv (\sigma_w \sqrt{f})^{-1} (1 + \lambda)$, where λ is a non-trivial eigenvalue of $\mathbf{J}_{\mathbf{w}}(\mathbf{x}^a)$ and $\mathbf{x}^a$ is an equilibrium with a fraction $f^a = f$ of active neurons. Main panel: cumulative distribution function of $|\tilde{\lambda}|$; Inset: full eigenspectrum. Black: theory; Colors: 3 equilibria randomly chosen from 3 different sample networks.

f , i.e., $q^{ab}(f) = Q(f) \delta_{ab} + q(f) (1 - \delta_{ab})$, where δ_{ab} is the Kronecker symbol. With this, we finally arrive at

$$\mathbb{E}[P_{\mathbf{w}}^n] \sim \int_0^1 df \exp[nN\Sigma_q(f)] \sim \exp[nN\Sigma_q(f_q^*)], \quad (7)$$

for $n \geq 2$; f_q^* is the value of f that maximizes Σ_q . Thus, $\Sigma_q^* = \Sigma_q(f_q^*)$ (see Eq. (3)). For $n = 1$, we obtain $\mathbb{E}[P_{\mathbf{w}}] \sim \exp[N\Sigma_a(f_a^*)]$ and f_a^* is the value of f that maximizes Σ_a . We provide explicit expressions for $\Sigma_q(f)$ and $\Sigma_a(f)$ in [11].

In Fig. 1(a), we plot Σ_q^* as a function of σ_w . For $\sigma_w \leq \sqrt{2}$, $\Sigma_q^* = 0$ and, therefore, there is a unique equilibrium. Using $\rho(z; f, \sigma_w)$, we compute its instability index, $\alpha(f_q^*)$ [11]. The instability index of an equilibrium $\mathbf{x}^a$ is the fraction of eigenvalues of $\mathbf{J}_{\mathbf{w}}(\mathbf{x}^a)$ with positive real part. Thus, $\alpha(f_q^*)N$ is the dimension of the unstable manifold of the equilibrium [14]. We find $\alpha(f_q^*) = 0$ (see Fig. 1(b))

and, therefore, the equilibrium is stable. Thus, there is a unique, stable equilibrium for $\sigma_w \leq \sqrt{2}$ ($N \rightarrow \infty$).

For $\sigma_w > \sqrt{2}$, $\Sigma_q^* > 0$ and increases with σ_w ; hence, there is typically an exponentially large (in N) number of equilibria. These equilibria have $\alpha(f_q^*) > 0$, indicating that they are saddles (i.e., linearly unstable). Interestingly, their unstable manifolds are ‘low-dimensional’, i.e., their fractional dimension, given by $\alpha(f_q^*)$, is $\ll 1$. For instance, at $\sigma_w = 2.5$, less than 4% of the eigenvalues have positive real parts (see Fig. 1(b)). There exists also an exponential number of equilibria with $f \neq f_q^*$. In fact, $\Sigma_q(f) > 0$ over a range of f values, and this range also increases with σ_w (Fig. 1(c)). However, these equilibria are all saddles; there are no stable equilibria for $\sigma_w > \sqrt{2}$ ($N \rightarrow \infty$). Thus, in our model, the transition to chaos is also accompanied by the sudden appearance of numerous, unstable equilibria [3, 5, 7].

We compare Σ_a^* with Σ_q^* . For $\sigma_w \leq \sqrt{2}$, $\Sigma_a^* = \Sigma_q^*$, and the maximum occurs for the same value of f , i.e., $f_a^* = f_q^*$, while for $\sigma_w > \sqrt{2}$, $\Sigma_a^* > \Sigma_q^*$ (Fig. 1(a)). In this case, the corresponding maxima occur for $f_a^* > f_q^*$ (Fig. 1(c)). This indicates the presence of exponentially large, but exponentially rare, fluctuations in P_w across samples. Thus, Σ_a^* is not an accurate estimate of the typical number of equilibria. Interestingly, stable equilibria never appear as a result of these large fluctuations. Indeed, $\alpha(f) > 0$ whenever $\Sigma_a(f) > 0$ [11]. We note that the difference between Σ_a^* and Σ_q^* is quantitatively small in the range of σ_w that we have explored.

We seek to validate our theory numerically by searching the equilibria of sample networks with a root-finding algorithm (Newton’s method with backtracking). As we explain in [11], in our case (i.e., $\phi(x) = x\Theta(x)$) we can unequivocally decide whether the ‘equilibrium’ returned by the algorithm is a *true* equilibrium of Eq. (1). Extensive search (as described in [11]) is feasible only for relatively small networks ($N \leq 200$).

We start by computing estimates of Σ_q^* as a function of σ_w from the equilibria found numerically. The resulting estimates are plotted in Fig. 1(a) on top of the theoretical prediction. There is a fairly good agreement with the theory, despite the small values of N . In Fig. 1(d), we compare instead the spectral distribution of $\mathbf{J}_w(\mathbf{x}^a)$, where $\mathbf{x}^a$ is an equilibrium, with the theoretical prediction. Again, there is a very good agreement. We see this also in Fig. 1(b), where we compare the instability index, averaged over all the equilibria $\mathbf{x}^a$, computed by numerically diagonalizing $\mathbf{J}_w(\mathbf{x}^a)$, with the average instability index computed from the theory, which is simply $\alpha(f_q^*)$ for $N \rightarrow \infty$.

The overlaps $q^{ab}(f)$ encode information about the spatial organization of the equilibria. Let us consider the equilibria $\mathbf{x}^a$ with the same $f^a = f$. The corresponding ϕ^a ’s lay inside a spherical cap [15] on the N -sphere of radius $\sqrt{NQ(f)}$ centered at the origin (in the non-negative orthant); see Eq. (6). The pole of this spherical

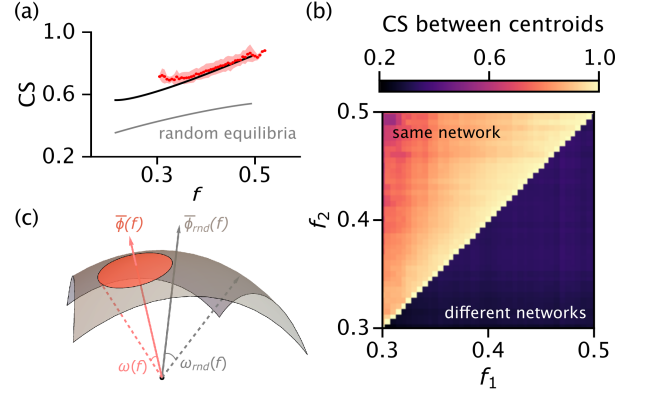


FIG. 2. (a) Cosine similarity (CS) between network’s equilibria and their centroid as a function of f . Full line: theory; Red dots: estimates from the equilibria numerically found (shaded area is $\pm$ SD across networks); Gray line: CS between random equilibria and their centroid. Parameters: $\sigma_w = 2.5$ and $N = 200$. (b) CS between centroids at different f_1 and f_2 , both in the same network (above diagonal), or in different networks (below diagonal). Parameters as in (a). (c) Cartoon illustrating the spatial organization of actual and random equilibria.

cap is given by the intersection of the N -sphere and the straight line defined by the ‘centroid’, $\bar{\phi}(f)$, that is, the average of the ϕ^a ’s. The polar angle, $\omega(f)$, satisfies $\cos \omega(f) = \sqrt{q(f)/Q(f)}$. This is illustrated in Fig. 2(a), where we plot, as a function of f , the CS between the equilibria numerically found and their centroid (red dots) together with the theoretical prediction, i.e., $\cos \omega(f)$ (black line) for $\sigma_w = 2.5$. As can be seen, there is good agreement.

This geometrical organization is natural in high-dimensional spaces ($N \rightarrow \infty$). To see this, consider ‘random’ equilibria, i.e., random vectors whose components are independently and identically distributed such that: (i) components are positive with probability f and 0 otherwise; (ii) their average is 1, as the average ϕ_i^a over equilibria and neurons; (iii) their variance is $Q(f) - 1$. It is easy to verify that also the random equilibria with a given f lay inside a spherical cap on the N -sphere of radius $\sqrt{NQ(f)}$, in the non-negative orthant [11]. The centroid, defining the pole, is the vector with all components equal to 1; the polar angle, $\omega_{\text{rnd}}(f)$, satisfies $\cos \omega_{\text{rnd}}(f) = \sqrt{1/Q(f)}$ (gray line in Fig. 2(a); for $\sigma_w = 2.5$).

In the whole range of $\sigma_w > \sqrt{2}$ that we have explored, it is always $q(f) > 1$ for all f such that $\Sigma_q(f) > 0$ and, therefore, $\omega_{\text{rnd}}(f) > \omega(f)$; compare the black and gray lines in Fig. 2(a). Hence, for any f , the cap containing the equilibria is smaller than the cap containing the corresponding random equilibria (see Fig. 2(c)). This is the geometrical manifestation of the correlations between equilibria. These originate from the fact that ϕ^a ’s are all equilibria of the *same* sample network. In fact, x_i^a and x_i^b ,

and hence ϕ_i^a and ϕ_i^b , (i.e., same neuron, different equilibria), are correlated because of the *quenched* couplings w_{ij} 's [16].

According to the above argument, we expect correlations also between equilibria with different values of f , i.e., f_1 and f_2 . A simple way to quantify these correlations is to compute the CS between the corresponding centroids, $\bar{\phi}(f_1)$ and $\bar{\phi}(f_2)$. In fact, this CS encodes the average overlap, $q(f_1, f_2)$, between equilibria ϕ^a with $f^a = f_1$ and equilibria ϕ^b with $f^b = f_2$, i.e.,

$$\cos \bar{\omega}(f_1, f_2) = \frac{q(f_1, f_2)}{\sqrt{q(f_1)q(f_2)}}. \quad (8)$$

For random equilibria, $q(f_1, f_2) = 1$. We have estimated the CS between $\bar{\phi}(f_1)$ and $\bar{\phi}(f_2)$ using the equilibria found numerically; the results are illustrated in Fig. 2(b). As can be seen, the estimated CSs are significantly larger than those expected if equilibria at f_1 and f_2 were random; equilibria are strongly correlated for each (f_1, f_2) . Geometrically, this implies that equilibria ‘occupy’ in the phase space a much smaller volume than random ones [17]. We note that the strong correlations are not inherited from correlations in the synaptic connectivity, in any obvious way at least, since the synaptic connectivity is completely random.

Finally, we study the network dynamics using DMFT [11]. In steady states, DMFT determines self-consistently the autocorrelation function

$$C(\tau) \equiv \lim_{N \rightarrow \infty} \frac{1}{N} \left(\tilde{\phi}^t \cdot \tilde{\phi}^{t+\tau} \right) \quad (9)$$

with $\tilde{\phi}^t \equiv (\phi(x_i(t)))$ where the $x_i(t)$'s are solutions of Eq. (1) (compare with Eq. (6)). For $N \rightarrow \infty$, $C(\tau)$ only depends on σ_w .

For $\sigma_w \leq \sqrt{2}$, $C(\tau) = Q_S$ is the only solution of DMFT. It corresponds to a unique, stable equilibrium and $Q_S = Q(f_q^*)$. For $\sigma_w > \sqrt{2}$, this solution still exists. However, there is only one Q_S , as before, and it is different from $Q(f_q^*)$ now. Thus, DMFT does not reveal all the equilibria of Eq. (1) (i.e., the ones with $Q \neq Q_S$) and those revealed are *not* the typical ones. For $\sigma_w > \sqrt{2}$, there are additional solutions which do not correspond to equilibria, i.e., $C(\tau)$ is not constant. We only consider the stable solution with $\lim_{\tau \rightarrow \infty} C(\tau) = q_D$, which describes the chaotic dynamics of Eq. (1) for $\sigma_w > \sqrt{2}$ (starting from random initial conditions).

We use the geometrical construction described above for the equilibria to ‘visualize’ the chaotic dynamics. To that end, we find it convenient to define $Q_D \equiv C(0)$. Clearly, $Q_D > q_D$. Then, the $\tilde{\phi}^t$'s lay inside a cap on the N -sphere of radius $\sqrt{N}Q_D$ centered at the origin, again in the non-negative orthant. The pole of the cap is given by the intersection of this N -sphere and the straight line defined by the centroid $\langle \tilde{\phi} \rangle$, i.e., the average over time of $\tilde{\phi}^t$. The polar angle satisfies $\cos \omega_D = \sqrt{q_D/Q_D}$. This is

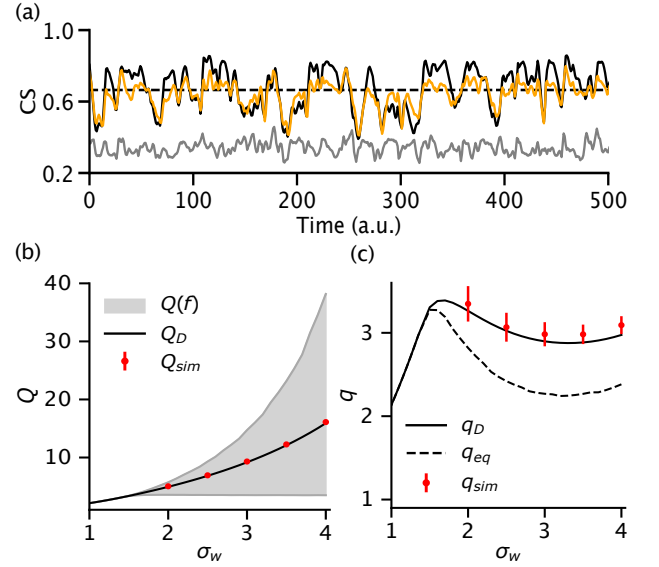


FIG. 3. (a) Cosine similarity between $\tilde{\phi}^t$ and $\langle \tilde{\phi} \rangle$ (black) and between $\tilde{\phi}^t$ and $\bar{\phi}$ (orange), when all vectors are computed in the *same* sample network ($\sigma_w = 2.5$, $N = 200$). For comparison, we also plot the cosine similarity between $\tilde{\phi}^t$ and $\bar{\phi}$, when these vectors are computed in *different* sample networks (gray). (b) Q_D as a function of σ_w . Red dots with errorbars (SD) represent numerical estimates ($n = 50$ samples with $N = 5000$). The gray area represents the range of $Q(f)$ over the corresponding equilibria. (c) q_D (full) and q_{eq} (dashed) as a function of σ_w . Red dots as in (b).

illustrated in Fig. 3(a), where we plot the time course of the CS between $\tilde{\phi}^t$ and $\langle \tilde{\phi} \rangle$.

To locate the ‘dynamical’ cap (i.e., the cap that contains the attractor) with respect to the equilibria, we also plot in Fig. 3(a) the CS between $\tilde{\phi}^t$ and $\bar{\phi}$, defined as the average of *all* the equilibria in the network (note that $\bar{\phi} \rightarrow \bar{\phi}(f_q^*)$ for $N \rightarrow \infty$). As can be seen, the two CSs are highly correlated, indicating that $\langle \tilde{\phi} \rangle$ and $\bar{\phi}$ are (almost) collinear. Does the dynamical cap lay inside the region that contains the equilibria?

To answer this question, we begin by determining whether there exists an f_{eq} such that $Q(f_{eq}) = Q_D$. Clearly, if such an f_{eq} does not exist, then the dynamical cap is outside. It turns out that a unique f_{eq} always exists. This is illustrated in Fig. 3(b), where we plot Q_D as a function of σ_w , together with the largest, Q_{max} , and the smallest, Q_{min} , value of Q over the equilibria with the same σ_w . As can be seen, $Q_{min} < Q_D < Q_{max}$ in the whole range of $\sigma_w > \sqrt{2}$ that we have explored. There is a unique f_{eq} because $Q(f)$ is a strictly monotonous decreasing function of f [11]. We note that it is always $Q_D > Q(f_q^*)$, and hence $f_{eq} < f_q^*$, for $\sigma_w > \sqrt{2}$ (not shown).

Next, using f_{eq} , we determine $q_{eq} \equiv q(f_{eq})$ to be compared with q_D . This is illustrated in Fig. 3(c), where we plot both q_D and q_{eq} as a function of σ_w . As can be seen,

in the whole range of $\sigma_w > \sqrt{2}$ that we have explored, it is $q_D > q_{eq}$. The dynamical cap lays always inside the region containing the equilibria, at a distance from the origin determined by Q_D . Thus, as already noticed for other models [6, 7], also in our case the dynamics occurs ‘far’ from the most numerous equilibria, because $Q_D > Q(f_q^*)$. Nevertheless, the dynamics always occurs *inside* a cap containing an exponential number of equilibria; in this sense, the dynamics occurs ‘close’ to specific equilibria, i.e., those with $f = f_{eq}$.

Our results provide a simple, geometric interpretation of a puzzling observation about the dynamics of large neuronal networks. In these networks, pair-wise correlations are vanishingly small, i.e., they are $\sim 1/\sqrt{N}$ and zero on average for $N \rightarrow \infty$. This seems to suggest that the degrees of freedom are essentially independent. Yet, the network’s states explored by the dynamics are completely dominated by correlation effects [2, 6] (see [18] for a general discussion). This is simply a consequence of the fact that the dynamics is restricted within a (very) small region of the phase space, as we have shown. This region appears to be determined by the geometric organization of the equilibria.

For $N \rightarrow \infty$, all points on the attractor have the same fraction of active neurons, $f = f_D$, which can be computed from DMFT. The (fractional) dimension of their unstable manifold, then, is simply given by $\alpha(f_D)$, i.e., the instability index of any point on the attractor. In particular, $\alpha(f_D)$ is the fraction of positive Lyapunov exponents and, as such, it provides an easily accessible quantification of the attractor dimension [19]. As we have shown, it is always $f_D < f_q^*$; hence, the attractor dimension is smaller than $\alpha(f_q^*)$, which can be computed *without* solving DMFT. We note that, in our model, even the ‘most unstable’ equilibrium (i.e., the equilibrium with the largest number of unstable directions) has a very small instability index.

We conjecture that, in the mean-field limit, the attractor dimension cannot exceed the largest instability index. This would be the case, for instance, if the attractor is contained in the closure of the unstable manifolds of some equilibria [14]. Our conjecture can be disproved by numerically evaluating Lyapunov spectra [19]. If, on the other hand, our conjecture is true, then low-dimensional dynamics is a generic property of large neuronal networks. Indeed, equilibria with small instability indexes are not a peculiarity of our model; rather, this appears to be a ‘typical’ feature of a large class of mean-field models [4]. Consistently, different measures of ‘dimensionality’ reveal that the dynamics in these models is low-dimensional [2, 6, 19].

[†] gianluigi.mongillo@gmail.com

- [1] A. Crisanti and H. Sompolinsky, Path integral approach to random neural networks, *Physical Review E* **98**, 062120 (2018).
- [2] D. G. Clark, L. Abbott, and A. Litwin-Kumar, Dimension of activity in random neural networks, *Physical Review Letters* **131**, 118401 (2023).
- [3] G. Wainrib and J. Touboul, Topological and dynamical complexity of random neural networks, *Physical Review Letters* **110**, 118101 (2013).
- [4] G. Ben Arous, Y. V. Fyodorov, and B. A. Khoruzhenko, Counting equilibria of large complex systems by instability index, *Proceedings of the National Academy of Sciences* **118**, e2023719118 (2021).
- [5] V. Ros, F. Roy, G. Biroli, G. Bunin, and A. M. Turner, Generalized lotka-volterra equations with random, non-reciprocal interactions: The typical number of equilibria, *Physical Review Letters* **130**, 257401 (2023).
- [6] J. Stubenrauch, C. Keup, A. C. Kurth, M. Helias, and A. van Meegen, Fixed point geometry in chaotic neural networks, *Physical Review Research* **7**, 023203 (2025).
- [7] S. J. Fournier, A. Pocco, V. Ros, and P. Urbani, Non-reciprocal interactions and high-dimensional chaos: comparing dynamics and statistics of equilibria in a solvable model, *arXiv preprint arXiv:2503.20908* (2025).
- [8] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Chaos in random neural networks, *Physical review letters* **61**, 259 (1988).
- [9] Our network operates in the so-called balanced regime [11]; this fixes the average level of activity, i.e., $\frac{1}{N} \sum_j \phi_j = 1 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$.
- [10] As an historical aside, we note that neither Kac nor Rice ever wrote such a formula, let alone with the purpose of counting the solutions of a large system of non-linear equations. Instead, for that precise purpose, Eq. (2) and an accompanying theory first appear in [20] to the best of our knowledge.
- [11] See Supplemental Material at [URL-will-be-inserted-by-publisher] for details.
- [12] Y. Ahmadian, F. Fumarola, and K. D. Miller, Properties of networks with partially structured and partially random connectivity, *Physical Review E* **91**, 012820 (2015).
- [13] M. Mezard and A. Montanari, *Information, physics, and computation* (Oxford University Press, 2009).
- [14] J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, *Reviews of modern physics* **57**, 617 (1985).
- [15] For $N \rightarrow \infty$, the volume of the cap, and hence the equilibria, will be concentrated in the rim.
- [16] Correlations between x_i^a and x_j^a (i.e., different neurons, same equilibrium) are negligible because of the mean-field nature of the model.
- [17] Clearly, equilibria, random or otherwise, fill up zero volume. What we mean is that the actual equilibria can be enclosed in a volume that is significantly smaller than the volume needed to enclose the corresponding random equilibria.
- [18] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature* **440**, 1007 (2006).
- [19] R. Engelken, F. Wolf, and L. F. Abbott, Lyapunov spectra of chaotic recurrent neural networks, *Physical Review*

* giancarlo.lacamera@stonybrook.edu

- Research **5**, 043044 (2023).
- [20] A. J. Bray and M. A. Moore, Metastable states in spin glasses, *Journal of Physics C: Solid State Physics* **13**, L469 (1980).

Supplemental Material for “On the relationship between equilibria and dynamics in large, random neuronal networks”

Xiaoyu Yang,^{1,2,3} Giancarlo La Camera,^{2,3,4,5,6,*} and Gianluigi Mongillo^{7,8,9,†}

¹*Graduate Program in Physics and Astronomy,
Stony Brook University, Stony Brook, NY, USA*

²*Department of Neurobiology and Behavior, Stony Brook University, Stony Brook, NY, USA*

³*Center for Neural Circuit Dynamics, Stony Brook University, Stony Brook, NY, USA*

⁴*AI Innovation Institute, Stony Brook University, Stony Brook, NY, USA*

⁵*Center for Advanced Computational Science,
Stony Brook University, Stony Brook, NY, USA*

⁶*Graduate Programs in Neuroscience, Stony Brook University, Stony Brook, NY, USA*

⁷*Sorbonne Université, INSERM, CNRS, Institut de la Vision, F-75012 Paris, France*

⁸*Centre National de la Recherche Scientifique, Paris, France*

⁹*School of Natural Sciences, Institute for Advanced Study, Princeton, NJ, USA.*

CONTENTS

I	Counting equilibria using the replica method	2
A	Preliminaries	2
A.1	Instability index	3
B	Average over quenched disorders	4
C	Replica symmetric ansatz	5
D	Quenched complexity	6
D.1	Saddle-point equations for the single-equilibrium solution ($Q = q$)	7
D.2	Saddle-point equations for the multiple-equilibria solution ($Q > q$)	7
D.3	Finite-size correction for the quenched complexity	9
E	Annealed complexity	9
E.1	Finite-size correction for the annealed complexity	11
II	Geometry of the equilibria	12
III	Dynamical mean-field theory	15
IV	Numerical simulations	16
A	Newton’s method with backtracking	16
A.1	Determining the number of initial conditions	17
A.2	Identifying true and unique equilibria	17
A.3	Numerical estimation of complexity	17
B	Numerical estimation of order parameters	18
	References	20

* giancarlo.lacamera@stonybrook.edu

† gianluigi.mongillo@gmail.com

I. COUNTING EQUILIBRIA USING THE REPLICA METHOD

A. Preliminaries

We consider a neuronal network model of N all-to-all connected, threshold-linear units, where the dynamics of each neuron's state x_i evolve according to

$$\dot{x}_i = -x_i + \sqrt{N} + \frac{1}{\sqrt{N}} \sum_{j=1}^N w_{ij} \phi(x_j). \quad (\text{S-1})$$

Here, $\phi(x) = x\Theta(x)$ with $\Theta(x)$ denoting the Heaviside step function. The synaptic efficacies w_{ij} are taken as i.i.d. Gaussian variables with mean $\mathbb{E}[w_{ij}] = -1$ and variance $\text{Var}[w_{ij}] = \sigma_w^2$. The system described by Eq. (S-1) is known to exhibit a *paradigmatic transition* between a unique, stable equilibrium and a chaotic attractor at $\sigma_w^{(c)} = \sqrt{2}$ [1]. In this work, we focus on the chaotic regime where $\sigma_w > \sqrt{2}$.

For a given realization of the synaptic matrix $\mathbf{w}$, the total number of equilibria of Eq. (S-1), denoted as $P_{\mathbf{w}}$, can be computed using the Kac-Rice formula

$$P_{\mathbf{w}} = \int_{-\infty}^{\infty} \left(\prod_{i=1}^N dx_i \right) \prod_{i=1}^N \delta \left(-x_i + \sqrt{N} + \frac{1}{\sqrt{N}} \sum_{j=1}^N w_{ij} \phi_j \right) |\det \mathbf{J}_{\mathbf{w}}(\mathbf{x})|, \quad (\text{S-2})$$

where $\mathbf{J}_{\mathbf{w}}(\mathbf{x})$ is the Jacobian matrix of the dynamical system evaluated at $\mathbf{x}$. $P_{\mathbf{w}}$ is a random variable that fluctuates across realizations of $\mathbf{w}$. However, its logarithm per site, $N^{-1} \log P_{\mathbf{w}}$, is expected to be a *self-averaging* quantity in the limit $N \rightarrow \infty$. This defines the *quenched complexity*

$$\Sigma_q^* = \lim_{N \rightarrow \infty} \frac{1}{N} \log P_{\mathbf{w}} = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log P_{\mathbf{w}}], \quad (\text{S-3})$$

where $\mathbb{E}[\cdot]$ denotes the average over w_{ij} . The quenched complexity Σ_q^* describes the asymptotic behavior of the typical value of $P_{\mathbf{w}}$. Alternatively, the asymptotic behavior of $\mathbb{E}[P_{\mathbf{w}}]$ defines the *annealed complexity*

$$\Sigma_a^* = \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[P_{\mathbf{w}}]. \quad (\text{S-4})$$

Evaluating Σ_q^* requires calculating the average of logarithms, which is typically a challenging task. To compute the quenched complexity, we use the replica method by first evaluating the average of $P_{\mathbf{w}}^n$ for integer n and then taking the limit $n \rightarrow 0$. This yields

$$\Sigma_q^* = \lim_{N \rightarrow \infty} \lim_{n \rightarrow 0} \frac{1}{Nn} (\mathbb{E}[P_{\mathbf{w}}^n] - 1) = \lim_{n \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{Nn} \log \mathbb{E}[P_{\mathbf{w}}^n], \quad (\text{S-5})$$

where we adopted the common assumption that the two limits, $n \rightarrow 0$ and $N \rightarrow \infty$, can be exchanged safely.

Using the Fourier transforms of the δ -functions, the replicated Kac-Rice formula is given by

$$P_{\mathbf{w}}^n = \int_{-\infty}^{\infty} \left(\prod_{a=1}^n \prod_{i=1}^N \frac{dx_i^a d\tilde{x}_i^a}{2\pi} \right) \exp \left[\sum_{a=1}^n \sum_{i=1}^N i\tilde{x}_i^a (x_i^a - \sqrt{N} - \frac{1}{\sqrt{N}} \sum_{j=1}^N w_{ij} \phi_j) \right] \times \prod_{a=1}^n |\det \mathbf{J}_{\mathbf{w}}(\mathbf{x}^a)|, \quad (\text{S-6})$$

We parameterize the synaptic weights as $w_{ij} = -1 + \sigma_w z_{ij}$, where $z_{ij} \sim \mathcal{N}(0, 1)$ are standard Gaussian variables. We define $\mu^a = \sqrt{N} (1 - \frac{1}{N} \sum_{i=1}^N \phi_i^a)$ as the mean input to replica a . After collecting terms with

same coefficient z_{ij} , we arrive at the following expression

$$P_{\mathbf{w}}^n = \int_{-\infty}^{\infty} \left(\prod_{a=1}^n \prod_{i=1}^N \frac{dx_i^a d\tilde{x}_i^a}{2\pi} \right) \exp \left[\sum_{a=1}^n \sum_{i=1}^N i\tilde{x}_i^a (x_i^a - \mu^a) + \sum_{ij} z_{ij} \left(\frac{\sigma_w}{\sqrt{N}} \sum_{a=1}^n i\tilde{x}_i^a \phi_j^a \right) \right] \times \prod_{a=1}^n |\det \mathbf{J}_{\mathbf{w}}(\mathbf{x}^a)|. \quad (\text{S-7})$$

The Jacobian determinants in the last term are evaluated at different states $\mathbf{x}^a$ but for the same realization $\mathbf{w}$, hence, they are correlated with each other and with the delta functions (in the Fourier form). Computing the expectation of such a product is hard and often impractical. A previous study, however, suggested that, in the limit $N \rightarrow \infty$, the empirical spectral density of the Jacobian of certain disordered systems is *self-averaging* [2]. Their theory applies to our model network in Eq. (S-1). It turns out that, for a state $\mathbf{x}$ with Nf active neurons, $\mathbf{J}_{\mathbf{w}}(\mathbf{x})$ has one real eigenvalue of $-\sqrt{N}$, $(1-f)N$ eigenvalues of -1 , and the remaining fN eigenvalues are uniformly distributed within the disc of radius $R = \sigma_w \sqrt{f}$ centered at $-1 + i0$. We denote this nontrivial distribution as $\rho(z; f, \sigma_w)$.

The absolute value of the Jacobian determinant, $|\det \mathbf{J}(\mathbf{x})| = \prod_{i=1}^N |\lambda_i|$, is governed by the fN nontrivial eigenvalues. For large N , we approximate the determinant using the empirical spectral density $\rho(z; f, \sigma_w)$

$$|\det \mathbf{J}(\mathbf{x})| = \exp \left[Nf \cdot \frac{1}{Nf} \sum_{i=1}^N \ln |\lambda_i| \right] \stackrel{N \rightarrow \infty}{\approx} \exp \left[Nf \int dz \rho(z; f, \sigma_w) \log |z| \right]. \quad (\text{S-8})$$

To leading order, this simplifies to

$$|\det \mathbf{J}(\mathbf{x})| \sim \exp [N\zeta(f, \sigma_w)], \quad (\text{S-9})$$

where $\zeta(f, \sigma) \equiv f \int dz \rho(z; f, \sigma_w) \log |z|$. We assume that the spectral density $\rho(z; f, \sigma_w)$ is weakly correlated with the delta functions (which impose the equilibrium condition). This allows us to use the asymptotic relation in Eq. (S-9) also when $\mathbf{x}$ is an equilibrium of the system. Using the residue theorem, one can derive the following analytical expression for $\zeta(f, \sigma_w)$:

$$\zeta(f, \sigma_w) = \begin{cases} \frac{1}{2\sigma_w^2} + \frac{f}{2} \ln(\sigma_w^2 f) - \frac{f}{2} & \sigma_w \sqrt{f} > 1, \\ 0 & \sigma_w \sqrt{f} \leq 1. \end{cases} \quad (\text{S-10})$$

Furthermore, we make the assumption that the Jacobian determinants corresponding to different replicas are also weakly correlated for large N . This implies that the product of the determinants in Eq. (S-7) has a leading order of

$$\prod_{a=1}^n |\det \mathbf{J}_{\mathbf{w}}(\mathbf{x}^a)| \sim \exp \left[N \sum_{a=1}^n \zeta(f_a, \sigma_w) \right]. \quad (\text{S-11})$$

We note that now the only term that depends on $\mathbf{w}$ in Eq. (S-7) is $\exp \left[\sum_{ij} z_{ij} \left(\frac{\sigma_w}{\sqrt{N}} \sum_{a=1}^n i\tilde{x}_i^a \phi_j^a \right) \right]$, which is linear in the z_{ij} 's.

A.1. Instability index

The stability of an equilibrium $\mathbf{x}^a$ is determined by its corresponding Jacobian matrix $\mathbf{J}(\mathbf{x}^a)$: the equilibrium is stable if all eigenvalues have non-positive real parts and unstable otherwise. We quantify this stability

using the *instability index* α ($0 \leq \alpha \leq 1$), defined as the fraction of eigenvalues of $\mathbf{J}(\mathbf{x}^a)$ with positive real parts (i.e., the fractional dimension of the unstable manifold). For large N , the instability index α depends only on the spectral density $\rho(z; f, \sigma_w)$ and takes the following form:

$$\alpha(f) = \begin{cases} \frac{f}{\pi} \left(\arccos \sqrt{\frac{f_{\text{st}}}{f}} - \sqrt{\frac{f_{\text{st}}}{f} \left(1 - \frac{f_{\text{st}}}{f} \right)} \right) & \text{if } f > f_{\text{st}} \\ 0 & \text{if } f \leq f_{\text{st}} \end{cases} \quad (\text{S-12})$$

Here, $f_{\text{st}} = \sigma_w^{-2}$ denotes the stability threshold: equilibria with $f \leq f_{\text{st}}$ are stable ($\alpha = 0$), while those with $f > f_{\text{st}}$ are unstable ($\alpha > 0$).

B. Average over quenched disorders

The average over Gaussian variables z_{ij} can be performed straightforwardly

$$\left\langle \exp \left[\sum_{ij} z_{ij} \left(\frac{\sigma_w}{\sqrt{N}} \sum_{a=1}^n i \tilde{x}_i^a \phi_j^a \right) \right] \right\rangle_{z_{ij}} = \exp \left[-\frac{\sigma_w^2}{2} \sum_{a,b=1}^n \sum_{i=1}^N \tilde{x}_i^a \tilde{x}_i^b \underbrace{\left(\frac{1}{N} \sum_{j=1}^N \phi_j^a \phi_j^b \right)}_{q^{ab}} \right]. \quad (\text{S-13})$$

The *overlaps*, $q^{ab} \equiv \frac{1}{N} \sum_{i=1}^N \phi_i^a \phi_i^b$, measure the scalar products between two equilibria ϕ^a, ϕ^b with the same f . To decouple different sites, we introduce a set of *order parameters* $\{q_{ab}, \mu_a, f_a\}$ as follows

$$1 = \int dq_{ab} d\hat{q}_{ab} \exp \left[N \hat{q}_{ab} q_{ab} - \sum_{i=1}^N \hat{q}_{ab} \phi_i^a \phi_i^b \right], \quad (\text{S-14})$$

$$1 = \int df_a d\hat{f}_a \exp \left[N f_a \hat{f}_a - \hat{f}_a \sum_{i=1}^N \Theta_i^a \right], \quad (\text{S-15})$$

$$1 = \int d\mu_a d\hat{\mu}_a \exp \left[N \hat{\mu}_a - \hat{\mu}_a \sum_{i=1}^N \phi_i^a + \mathcal{O}(\sqrt{N}) \right]. \quad (\text{S-16})$$

Using these equations in Eq. (S-7), we obtain

$$\mathbb{E}[P_{\mathbf{w}}^n] = \int d\theta d\hat{\theta} \exp \left[N \Phi_n(\theta, \hat{\theta}) \right]. \quad (\text{S-17})$$

Here, $\theta = \{q_{ab}, \mu_a, f_a\}$ denotes the set of order parameters, and $\hat{\theta} = \{\hat{q}_{ab}, \hat{\mu}_a, \hat{f}_a\}$ their conjugate variables. The action $\Phi_n(\theta, \hat{\theta})$ is given by

$$\Phi_n(\theta, \hat{\theta}) = \sum_{a \leq b} \hat{q}_{ab} q_{ab} + \sum_a \hat{\mu}_a + \sum_a f_a \hat{f}_a + \sum_a \zeta(f_a, \sigma_w) + \log \Omega_n(\theta, \hat{\theta}), \quad (\text{S-18})$$

with

$$\begin{aligned} \Omega_n(\theta, \hat{\theta}) &= \int_{-\infty}^{\infty} \left(\prod_{a=1}^n dx_a \right) \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T (\sigma_w^2 \mathbf{Q})^{-1} (\mathbf{x} - \boldsymbol{\mu}) - \frac{1}{2} \ln \det(2\pi \sigma_w^2 \mathbf{Q}) \right] \\ &\times \exp \left[-\sum_{a \leq b} \hat{q}_{ab} \phi_a \phi_b - \sum_a \hat{\mu}_a \phi_a - \sum_a \hat{f}_a \Theta_a \right]. \end{aligned} \quad (\text{S-19})$$

Here, $\mathbf{x}$ and $\boldsymbol{\mu}$ are n -dimensional vectors in the replica space (each component corresponds to a replica index $a = 1, \dots, n$). $\mathbf{Q}$ is the $n \times n$ matrix with elements q^{ab} .

C. Replica symmetric ansatz

To compute the integral in Eq. (S-19), we need to make an explicit assumption about the symmetry of the order parameters under replica permutations. We proceed with the simplest assumption, the *replica symmetric (RS) ansatz*, which posits that:

$$f_a = f, \quad (\text{S-20})$$

$$\mu_a = \mu, \quad (\text{S-21})$$

$$q_{ab} = \delta_{ab}Q + (1 - \delta_{ab})q, \quad (\text{S-22})$$

and similarly for the conjugate variables (the choice of the minus sign in $\hat{q}$ is for notational convenience)

$$\hat{f}_a = \hat{f}, \quad (\text{S-23})$$

$$\hat{\mu}_a = \hat{\mu}, \quad (\text{S-24})$$

$$\hat{q}_{ab} = \delta_{ab}\hat{Q} - (1 - \delta_{ab})\hat{q}. \quad (\text{S-25})$$

With the RS symmetry, the partition function $\Omega_n(\theta, \hat{\theta})$ can be written as

$$\begin{aligned} \Omega_n(\theta, \hat{\theta}) &\propto \int_{-\infty}^{\infty} \left(\prod_{a=1}^n dx_a \right) \exp \left[-\frac{n^2 B}{2} \left(\frac{1}{n} \sum_{a=1}^n x_a - \mu \right)^2 + \frac{n^2 \hat{q}}{2} \left(\frac{1}{n} \sum_{a=1}^n \phi_a \right)^2 \right] \\ &\times \prod_{a=1}^n \exp \left[-\frac{A - B}{2} (x_a - \mu)^2 - (\hat{Q} + \hat{q}/2) \phi_a^2 - \hat{\mu} \phi_a - \hat{f} \Theta_a \right], \end{aligned} \quad (\text{S-26})$$

where A, B denotes the diagonal and off-diagonal elements of $(\sigma_w^2 \mathbf{Q})^{-1}$, respectively

$$A = \frac{Q - q + (n-1)q}{\sigma_w^2(Q - q)[Q + (n-1)q]}, \quad (\text{S-27})$$

$$B = \frac{-q}{\sigma_w^2(Q - q)[Q + (n-1)q]}. \quad (\text{S-28})$$

We introduce two auxiliary variables y, p as follows

$$1 = \int_{-\infty}^{\infty} dy d\tilde{y} \exp \left[\frac{in\tilde{y}}{\sigma_w^2(Q - q)} \left(y - \frac{1}{n} \sum_{a=1}^n x_a \right) \right], \quad (\text{S-29})$$

$$1 = \int_{-\infty}^{\infty} dp d\tilde{p} \exp \left[in\tilde{p} \left(p - \frac{1}{n} \sum_{a=1}^n \phi_a \right) \right], \quad (\text{S-30})$$

and plug them into $\Omega_n(\theta, \hat{\theta})$

$$\begin{aligned} \Omega_n(\theta, \hat{\theta}) &\propto \int_{-\infty}^{\infty} dp d\tilde{p} \exp \left[\frac{1}{2} n^2 \hat{q} p^2 + i n \tilde{p} p \right] \\ &\times \int_{-\infty}^{\infty} dy d\tilde{y} \exp \left[-\frac{n^2 B}{2} (y - \mu)^2 + \frac{i n \tilde{y} (y - \mu)}{\sigma_w^2 (Q - q)} - \frac{n \tilde{y}^2}{2 \sigma_w^2 (Q - q)} \right] \\ &\times \prod_{a=1}^n \int_{-\infty}^{\infty} \frac{dx_a}{\sigma_w \sqrt{2\pi(Q-q)}} \exp \left[-\frac{(x_a - \mu + i \tilde{y})^2}{2 \sigma_w^2 (Q - q)} - (\hat{Q} + \hat{q}/2) \phi_a^2 - (\hat{\mu} + i \tilde{p}) \phi_a - \hat{f} \Theta_a \right] \end{aligned} \quad (\text{S-31})$$

Integrating out y and p leads to

$$\begin{aligned} \Omega_n(\theta, \hat{\theta}) &= \int_{-\infty}^{\infty} d\tilde{y} d\tilde{p} \exp \left[\frac{\tilde{y}^2}{2 \sigma_w^2 q} - \frac{1}{2} \ln(2\pi \sigma_w^2 q) + \frac{\tilde{p}^2}{2 \hat{q}} - \frac{1}{2} \ln(2\pi \hat{q}) \right] \\ &\times \prod_{a=1}^n \int_{-\infty}^{\infty} \frac{dx_a}{\sigma_w \sqrt{2\pi(Q-q)}} \exp \left[-\frac{(x_a - \mu + i \tilde{y})^2}{2 \sigma_w^2 (Q - q)} - (\hat{Q} + \hat{q}/2) \phi_a^2 - (\hat{\mu} + i \tilde{p}) \phi_a - \hat{f} \Theta_a \right]. \end{aligned} \quad (\text{S-32})$$

$\tilde{y}$, $\tilde{p}$ can be expressed in terms of a pair of independent, standard Gaussian variables ξ , t as follows

$$\tilde{y} = i \sigma_w \sqrt{\hat{q}} \xi, \quad (\text{S-33})$$

$$\tilde{p} = -i t \sqrt{\hat{q}}. \quad (\text{S-34})$$

Therefore, the partition function $\Omega_n(\theta, \hat{\theta})$ can be expressed as

$$\Omega_n(\theta, \hat{\theta}) = \int \mathcal{D}\xi \mathcal{D}t [Z(\xi, t)]^n, \quad (\text{S-35})$$

where

$$Z(\xi, t) = \int_{-\infty}^{\infty} \frac{dx}{\sigma_w \sqrt{2\pi(Q-q)}} \exp \left[-\frac{(x - \mu - \xi \sigma_w \sqrt{\hat{q}})^2}{2 \sigma_w^2 (Q - q)} - (\hat{Q} + \hat{q}/2) \phi^2 - (\hat{\mu} + t \sqrt{\hat{q}}) \phi - \hat{f} \Theta \right]. \quad (\text{S-36})$$

Using again the replica trick, we have

$$\lim_{n \rightarrow 0} \frac{1}{n} (\Omega_n - 1) = \int \mathcal{D}\xi \mathcal{D}t \log Z(\xi, t), \quad (\text{S-37})$$

thus

$$\lim_{n \rightarrow 0} \frac{1}{n} \log \Omega_n = \lim_{n \rightarrow 0} \frac{1}{n} \log \left(1 + n \cdot \frac{1}{n} (\Omega_n - 1) \right) = \int \mathcal{D}\xi \mathcal{D}t \log Z(\xi, t). \quad (\text{S-38})$$

D. Quenched complexity

Using the RS ansatz in Eq. (S-17), we obtain the complexity function

$$\mathbb{E}[P_{\mathbf{w}}^n] = \int d\theta d\hat{\theta} \exp \left[n N \Sigma_q(\theta, \hat{\theta}) \right] \sim \exp \left[n N \Sigma_q^* \right], \quad (\text{S-39})$$

with

$$\Sigma_q(\theta, \hat{\theta}) = \hat{Q}Q + \frac{1}{2}\hat{q}q + \hat{\mu} + \hat{f}f + \zeta(f, \sigma_w) + \int \mathcal{D}\xi \mathcal{D}t \log Z(\xi, t). \quad (\text{S-40})$$

The quenched complexity Σ_q^* corresponds to the extremum of $\Sigma_q(\theta, \hat{\theta})$, which is found by solving the saddle-point equations with respect to the order parameters.

D.1. Saddle-point equations for the single-equilibrium solution ($Q = q$)

We first consider the case $Q = q$, corresponding to the regime where the network has a unique equilibrium. In this setting, the complexity functional simplifies to

$$\Sigma_q(\theta, \hat{\theta}) = \hat{Q}(Q - \langle \phi^2 \rangle_\xi) + \hat{\mu}(1 - \langle \phi \rangle_\xi) + \hat{f}(f - \langle \Theta \rangle_\xi) + \frac{\hat{q}}{2}(q - \langle \phi^2 \rangle_\xi) + \zeta(f, \sigma_w), \quad (\text{S-41})$$

where $\phi = \phi(\mu + \xi\sigma_w\sqrt{Q})$ and $\Theta = \Theta(\mu + \xi\sigma_w\sqrt{Q})$ are shorthand notations. The saddle-point equations read

$$Q = \langle \phi^2 \rangle_\xi = \int \mathcal{D}\xi \phi^2(\mu + \xi\sigma_w\sqrt{Q}), \quad (\text{S-42})$$

$$1 = \langle \phi \rangle_\xi = \int \mathcal{D}\xi \phi(\mu + \xi\sigma_w\sqrt{Q}), \quad (\text{S-43})$$

$$f = \langle \Theta \rangle_\xi = \int \mathcal{D}\xi \Theta(\mu + \xi\sigma_w\sqrt{Q}). \quad (\text{S-44})$$

These equations admit a unique solution for any value of σ_w , and coincide with those derived using the cavity method [1]. While the $Q = q$ solution always exists, it maximizes the function $\Sigma_q(\theta, \hat{\theta})$ only for $\sigma_w \leq \sqrt{2}$. For $\sigma_w > \sqrt{2}$, the solution with $Q > q$ dominates, indicating the appearance of exponentially many equilibria, as we will show.

D.2. Saddle-point equations for the multiple-equilibria solution ($Q > q$)

When there are multiple equilibria, we define $\sigma = \sigma_w\sqrt{Q - q} > 0$, and rescale the auxiliary variables as

$$\hat{Q} \rightarrow \sigma^{-2}(\hat{Q} - \hat{q}/2), \quad (\text{S-45})$$

$$\hat{q} \rightarrow \sigma^{-2}\hat{q}, \quad (\text{S-46})$$

$$\hat{\mu} = \sigma^{-1}\hat{\mu}. \quad (\text{S-47})$$

The partition function becomes

$$Z(\xi, t) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{x}{\sigma} - \frac{\mu}{\sigma} - \xi \frac{\sigma_w\sqrt{q}}{\sigma} \right)^2 - \hat{Q}\phi(x/\sigma)^2 - (\hat{\mu} + t\sqrt{\hat{q}})\phi(x/\sigma) - \hat{f}\Theta(x/\sigma) \right], \quad (\text{S-48})$$

and the complexity functional reads

$$\Sigma_q(\theta, \hat{\theta}) = f\hat{f} + \sigma^{-1}\hat{\mu} + \sigma^{-2}Q\hat{Q} - \frac{\hat{q}}{2\sigma_w^2} + \zeta(f, \sigma_w) + \int \mathcal{D}\xi \mathcal{D}t \log Z(\xi, t). \quad (\text{S-49})$$

We introduce the following variables

$$m = \frac{\mu}{\sigma}, \quad (\text{S-50})$$

$$z = \frac{\sigma_w \sqrt{\hat{q}}}{\sigma}. \quad (\text{S-51})$$

With these definitions, the partition function and the complexity functional become

$$Z(\xi, t) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x - m - \xi z)^2 - \hat{Q}\phi^2(x) - (\hat{\mu} + t\sqrt{\hat{q}})\phi(x) - \hat{f}\Theta(x) \right], \quad (\text{S-52})$$

and

$$\Sigma_q(\theta, \hat{\theta}) = f\hat{f} + \sigma^{-1}\hat{\mu} + \frac{\hat{Q}}{\sigma_w^2}(1 + z^2) - \frac{\hat{q}}{2\sigma_w^2} + \zeta(f, \sigma_w) + \int \mathcal{D}\xi \mathcal{D}t \log Z(\xi, t). \quad (\text{S-53})$$

We define the normalized probability distribution $p(x|\xi, t)$ as

$$p(x|\xi, t) \propto \frac{dx}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x - m - \xi z)^2 - \hat{Q}\phi^2 - (\hat{\mu} + t\sqrt{\hat{q}})\phi - \hat{f}\Theta(x) \right], \quad (\text{S-54})$$

and denote integration over x , ξ and t with respect to this distribution as $\langle \cdot \rangle$, i.e.,

$$\langle O(x|\xi, t) \rangle = \int \mathcal{D}\xi \mathcal{D}t \int_{-\infty}^{\infty} dx p(x|\xi, t) O(x, \xi, t). \quad (\text{S-55})$$

The saddle-point equations are obtained by taking derivatives of Σ_q with respect to the rescaled variables $\theta_{new} = \{\hat{f}, \hat{Q}, \hat{\mu}, \hat{q}, m, \sigma, z\}$. For σ , $\hat{\mu}$, $\hat{f}$, we have

$$\hat{\mu} = 0. \quad (\text{S-56})$$

$$\sigma^{-1} = \langle \phi \rangle. \quad (\text{S-57})$$

$$f = \langle \Theta \rangle. \quad (\text{S-58})$$

For $\hat{Q}$, $\hat{q}$, m , z , the saddle-point equations read (where we used the fact that $\langle \xi \rangle = 0$ and $\langle \xi^2 \rangle = 1$)

$$\frac{1 + z^2}{\sigma_w^2} = \langle \phi^2 \rangle. \quad (\text{S-59})$$

$$\frac{\hat{q}}{\sigma_w^2} + \langle t\phi \rangle = 0. \quad (\text{S-60})$$

$$m - \langle x \rangle = 0. \quad (\text{S-61})$$

$$\frac{2\hat{Q}z}{\sigma_w^2} + \langle \xi x \rangle - z = 0. \quad (\text{S-62})$$

Eqs. (S-59-S-62) can be solved self-consistently for a given $\hat{f}$, then we substitute the solutions into Eqs. (S-57, S-58) to determine the remaining parameters.

D.3. Finite-size correction for the quenched complexity

We consider the leading-order correction to the quenched complexity by imposing an N -dependent balance condition

$$\langle \phi \rangle = 1 \quad \Rightarrow \quad \langle \phi \rangle = 1 - \frac{\mu}{\sqrt{N}}. \quad (\text{S-63})$$

This modification is equivalent to replacing the term $\hat{\mu} \cdot 1$ in the complexity function with $\hat{\mu}(1 - \frac{\mu}{\sqrt{N}})$. Expressed in terms of the rescaled variables, the complexity becomes

$$\Sigma_q(\theta, \hat{\theta}) = \Sigma_q^{(0)}(\theta, \hat{\theta}) - \frac{m}{\sqrt{N}} \hat{\mu} + \mathcal{O}(N^{-1}), \quad (\text{S-64})$$

where $\Sigma_q^{(0)}(\theta, \hat{\theta})$ denotes the asymptotic complexity. This correction affects only the saddle-point equation with respect to $\hat{\mu}$, which is modified to

$$\sigma^{-1} = \langle \phi \rangle + \frac{m}{\sqrt{N}}, \quad (\text{S-65})$$

while all other equations remain unchanged (the equation for m is unaffected, since the saddle-point conditions still enforce $\hat{\mu} = 0$).

Because both $\langle \phi \rangle$ and m can be determined independently of σ , their values coincide with those obtained in the limit $N \rightarrow \infty$. Using the definitions of m and σ , we obtain the following finite-size corrections

$$Q_N = Q_\infty \left(1 + \frac{\mu_\infty}{\sqrt{N}} \right)^{-2}, \quad (\text{S-66})$$

$$q_N = q_\infty \left(1 + \frac{\mu_\infty}{\sqrt{N}} \right)^{-2}, \quad (\text{S-67})$$

$$\mu_N = \mu_\infty \left(1 + \frac{\mu_\infty}{\sqrt{N}} \right)^{-1}. \quad (\text{S-68})$$

E. Annealed complexity

The average number of equilibria, $\mathbb{E}[P_{\mathbf{w}}]$, can be estimated by setting $n = 1$ in Eq. (S-19). Using the saddle-point approximation, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[P_{\mathbf{w}}^{n=1}] = \operatorname{argmax}_{\theta, \hat{\theta}} \Sigma_a(\theta, \hat{\theta}), \quad (\text{S-69})$$

where $\Sigma_a(\theta, \hat{\theta})$ is the corresponding functional for the annealed complexity

$$\Sigma_a(\theta, \hat{\theta}) = \hat{Q}Q + \hat{\mu} + \hat{f}f + \zeta(f, \sigma_w) + \ln \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma_w^2 Q}} \exp \left[-\frac{(x - \mu)^2}{2\sigma_w^2 Q} - \hat{Q}\phi^2 - \hat{\mu}\phi - \hat{f}\Theta \right]. \quad (\text{S-70})$$

The parameters θ and $\hat{\theta}$ can be determined by solving the corresponding saddle-point equations. As a first step, we evaluate the Gaussian integral for $x < 0$ in the last term

$$Z_- = \int_{-\infty}^0 \frac{dx}{\sqrt{2\pi\sigma_w^2 Q}} \exp \left[-\frac{(x-\mu)^2}{2\sigma_w^2 Q} \right] = \frac{1}{2} \text{erfc}(z), \quad z \equiv \frac{\mu}{\sqrt{2\sigma_w^2 Q}}. \quad (\text{S-71})$$

Similarly, for $x > 0$ we obtain

$$\begin{aligned} Z_+ &= \int_0^\infty \frac{dx}{\sqrt{2\pi\sigma_w^2 Q}} \exp \left[-\frac{(x-\mu)^2}{2\sigma_w^2 Q} - \hat{\mu}x - \hat{Q}x^2 - \hat{f} \right] \\ &= \int_0^\infty \frac{dx}{\sqrt{2\pi\sigma_w^2 Q}} \exp \left[-\beta(x-\alpha)^2 + \beta\alpha^2 - z^2 - \hat{f} \right] \\ &= \frac{\text{erfc}(y)}{2\sqrt{2\sigma_w^2 Q}\beta} \exp \left[y^2 - z^2 - \hat{f} \right], \quad y \equiv -\alpha\sqrt{\beta}, \end{aligned} \quad (\text{S-72})$$

where β and α are defined by completing the square in the exponent for $x > 0$

$$\beta = \frac{1}{2\sigma_w^2 Q} + \hat{Q}, \quad \alpha = \frac{1}{2\beta} \left(\frac{\mu}{\sigma_w^2 Q} - \hat{\mu} \right). \quad (\text{S-73})$$

The saddle-point equation with respect to $\hat{f}$ gives

$$f - \frac{Z_+}{Z_- + Z_+} = 0 \quad \Rightarrow \quad \frac{Z_+}{Z_-} = \frac{f}{1-f} = \frac{\text{erfc}(y)}{\text{erfc}(z)} \frac{\exp[y^2 - z^2 - \hat{f}]}{\sqrt{2\sigma_w^2 Q}\beta}. \quad (\text{S-74})$$

Hence, $\hat{f}$ is uniquely determined by the remaining parameters. Substituting the solution for $\hat{f}$ allows us to split the logarithm term

$$\begin{aligned} \ln(Z_- + Z_+) &= \ln \left[(1-f) \cdot \frac{Z_-}{1-f} + f \cdot \frac{Z_+}{f} \right] \\ &= (1-f) \ln Z_- + f \ln Z_+ - f \ln f - (1-f) \ln(1-f). \end{aligned} \quad (\text{S-75})$$

Substituting $\hat{Q} \rightarrow \beta - \frac{1}{2\sigma_w^2 Q}$ and $\hat{\mu} \rightarrow \frac{\sqrt{2}z}{\sigma_w^2 Q} + 2\sqrt{\beta}y$ yields the complexity function

$$\begin{aligned} \Sigma_a(y, z, \beta, Q) &= (1-f) \ln \text{erfc}(z) - fz^2 + \frac{2z}{\sqrt{2\sigma_w^2 Q}} + f \ln \text{erfc}(y) + fy^2 + 2\sqrt{\beta}y \\ &\quad + \beta Q - \frac{f}{2} \ln(2\sigma_w^2 Q\beta) + \zeta(f, \sigma_w) - \frac{1}{2\sigma_w^2} - f \ln f - (1-f) \ln(1-f) - \ln 2, \end{aligned} \quad (\text{S-76})$$

We introduce the function

$$\psi(x) = \frac{\exp(-x^2)}{\sqrt{\pi} \text{erfc}(x)} = -\frac{1}{2} \frac{d}{dx} \ln \text{erfc}(x), \quad (\text{S-77})$$

and the saddle-point equations with respect to z , y , Q and β are given by

$$fz + (1-f)\psi(z) = \frac{1}{\sqrt{2\sigma_w^2 Q}}, \quad (\text{S-78})$$

$$f(\psi(y) - y) = \sqrt{\beta}, \quad (\text{S-79})$$

$$-\frac{2z}{\sqrt{2\sigma_w^2 Q}} + 2\beta Q - f = 0, \quad (\text{S-80})$$

$$2\sqrt{\beta}y + 2\beta Q - f = 0, \quad (\text{S-81})$$

which can be summarized into one equation

$$f - 2\beta Q = 2fy(\psi(y) - y) = -2z(fz + (1 - f)\psi(z)). \quad (\text{S-82})$$

We observe that $2y(\psi(y) - y)$ is a monotonically increasing function with range $(-\infty, 1)$. Thus, for a given f and $\lambda \equiv 2\beta Q > 0$, y is uniquely determined as a function of λ ; we denote this solution by y_λ . From Eqs. (S-79, S-80), z_λ can also be computed as a function of λ

$$z_\lambda = -\sigma_w y_\lambda \sqrt{\lambda}. \quad (\text{S-83})$$

Substituting Eq. (S-83) into Eq. (S-82) yields the final equation for λ

$$f - \lambda + 2z_\lambda[fz_\lambda + (1 - f)\psi(z_\lambda)] = 0. \quad (\text{S-84})$$

E.1. Finite-size correction for the annealed complexity

The finite-size correction to the annealed complexity is obtained in the same way as for the quenched case. We replace $\hat{\mu}$ with $\hat{\mu}(1 - \frac{\mu}{\sqrt{N}})$ in Eq. (S-70) and obtain the following modified saddle-point equations

$$fz + (1 - f)\psi(z) + \frac{2z}{\sqrt{N}} + \frac{y}{\sqrt{N}}\sqrt{2\sigma_w^2 Q\beta} = \frac{1}{\sqrt{2\sigma_w^2 Q}}, \quad (\text{S-85})$$

$$f(\psi(y) - y) + \frac{z}{\sqrt{N}}\sqrt{2\sigma_w^2 Q\beta} = \sqrt{\beta}, \quad (\text{S-86})$$

$$2\beta Q - f - \frac{2z}{\sqrt{2\sigma_w^2 Q}} - \frac{2yz}{\sqrt{N}}\sqrt{2\sigma_w^2 Q\beta} = 0, \quad (\text{S-87})$$

$$2\beta Q - f + 2\sqrt{\beta}y - \frac{2yz}{\sqrt{N}}\sqrt{2\sigma_w^2 Q\beta} = 0. \quad (\text{S-88})$$

From the last two equations we obtain

$$\frac{z}{y} = -\sigma_w \sqrt{\lambda}, \quad (\text{S-89})$$

which implies

$$\hat{\mu} = \frac{\sqrt{2}z}{\sqrt{\sigma_w^2 Q}} + 2\sqrt{\beta}y = 0. \quad (\text{S-90})$$

The modified saddle-point equations yield the same solutions for y_λ and z_λ as functions of λ . Moreover, the annealed complexity itself remains unaffected by the finite-size correction (since $\hat{\mu} = 0$). As in the quenched case, finite-size effects appear only in the geometric parameters Q and μ , which scale according to Eqs. (S-66, S-68).

II. GEOMETRY OF THE EQUILIBRIA

Here we discuss the organization of the equilibria in the phase space, i.e., their geometry, as revealed by the replica-symmetric theory presented in the previous section. The theory determines a set of ‘order parameters’ as a function of the fraction of active neurons f (and the synaptic gain σ_w). Among these, $Q(f)$ and $q(f)$ have a simple geometrical interpretation. The distance of ϕ^a (with $f^a = f$) from the origin (i.e., the norm of ϕ^a) is given by

$$D^a(f) = \sqrt{\sum_{j=1}^N (\phi_j^a - 0)^2} = \sqrt{\frac{N}{N} \sum_{j=1}^N (\phi_j^a)^2} = \sqrt{NQ(f)} \equiv D(f) \quad (\text{S-91})$$

which does not depend on a in the limit $N \rightarrow \infty$. Therefore, all ϕ^a ’s lay on the hypersphere centered at the origin with radius $D(f)$, in the non-negative orthant, because $\phi_j^a \geq 0$ for all a and j by definition. The radius of the hypersphere decreases with f ; the larger f , the closer to the origin the corresponding ϕ^a ’s.

The angle, $\theta^{ab}(f)$, between two distinct ϕ^a and ϕ^b (i.e., $a \neq b$) with $f^a = f^b = f$, is given by

$$\cos \theta^{ab}(f) = \frac{\sum_{j=1}^N \phi_j^a \phi_j^b}{\sqrt{\sum_{j=1}^N (\phi_j^a)^2} \sqrt{\sum_{k=1}^N (\phi_k^b)^2}} = \frac{Nq(f)}{\sqrt{NQ(f)} \sqrt{NQ(f)}} = \frac{q(f)}{Q(f)} \equiv \cos \theta(f) \quad (\text{S-92})$$

which, again, does not depend on the specific solutions a and b chosen, but only on f . Therefore, any pair of distinct solutions, ϕ^a and ϕ^b , forms the same angle (at parity of f). This angle also decreases with f ; the larger f , the smaller the angle $\theta(f)$, and the larger the cosine similarity, i.e., $\cos \theta(f)$, between the corresponding ϕ^a ’s.

The fact that any pair of distinct solutions makes the same angle is not surprising in high-dimensional spaces; it is just a manifestation of the law of large numbers. To see this, let us consider the constraints that any solution ϕ^a with $f^a = f$ has to satisfy (in the limit $N \rightarrow \infty$). There are three constraints; they are

$$\phi_j^a \geq 0 \text{ for all } j, \quad (\text{S-93})$$

because $\phi(x) = x\Theta(x)$,

$$\frac{1}{N} \sum_i \phi_i^a = 1, \quad (\text{S-94})$$

which is a consequence of the balanced regime our model network operates in, and, finally,

$$\frac{1}{N} \sum_i (\phi_i^a)^2 = Q(f), \quad (\text{S-95})$$

which fixes the norm of the ϕ^a . What is the angle formed by two vectors, $\hat{\phi}^a$ and $\hat{\phi}^b$, whose elements are randomly (according to some distribution) and *independently* chosen such that the above constraints are satisfied? To answer the question, we evaluate their scalar product. This is given by

$$\hat{\phi}^a \cdot \hat{\phi}^b = \sum_i \hat{\phi}_i^a \hat{\phi}_i^b = N f^2 \langle \hat{\phi}_i^a \rangle_A \langle \hat{\phi}_i^b \rangle_A = N f^2 \frac{1}{f} \frac{1}{f} = N \quad (\text{S-96})$$

because the probability that both $\hat{\phi}_i^a$ and $\hat{\phi}_i^b$ are different from zero is f^2 , and the average of $\hat{\phi}_i^a$ restricted to the ‘active’ neurons, that we have denoted $\langle \hat{\phi}_i^a \rangle_A$ in the above equation, is $1/f$ (because of Eq. (S-94)). Thus, the angle they form, $\hat{\theta}(f)$, is given by

$$\cos \hat{\theta}(f) = \frac{\hat{\phi}^a \cdot \hat{\phi}^b}{\|\hat{\phi}^a\| \|\hat{\phi}^b\|} = \frac{N}{\sqrt{NQ(f)} \sqrt{NQ(f)}} = \frac{1}{Q(f)} \quad (\text{S-97})$$

and it is independent of $\hat{\phi}^a$ and $\hat{\phi}^b$. Comparing Eq. (S-97) with Eq. (S-92), and noting that $q(f) > 1$, we see that $\theta(f) < \hat{\theta}(f)$. In other words, actual equilibria are more similar, as measured by cosine similarity, than random vectors satisfying the same constraints, that is, actual equilibria are significantly correlated.

The origin of these correlations should be obvious; they result from the fact that ϕ^a ’s are all equilibria of the *same* sample network. In fact, x_i^a and x_i^b , and hence ϕ_i^a and ϕ_i^b (i.e., same neuron, different equilibria), are correlated because the w_{ij} ’s do not change with the equilibria (i.e., the w_{ij} ’s are quenched variables). These correlations are quantified by $q(f)$ as discussed above, and they are only accessible through the quenched theory; in the annealed theory there is no $q(f)$, because of the *direct* averaging over the w_{ij} , which effectively removes the correlations. Instead, correlations between ϕ_i^a and ϕ_j^a (i.e., same equilibrium, different neurons) are negligible in the limit $N \rightarrow \infty$, because of the mean-field nature of our model, both in the quenched and the annealed theory.

To better understand the spatial organization of the equilibria, let us consider a (very) large sample network. For any f such that $\Sigma_q(f) > 0$, we can define the *center-of-mass* of the corresponding ϕ^a ’s (i.e., with $f^a = f$) as

$$\bar{\phi}(f) = \frac{1}{P_{\mathbf{w}}(f)} \sum_{a|f^a=f} \phi^a \quad (\text{S-98})$$

where $P_{\mathbf{w}}(f)$ is the number of equilibria with $f^a = f$ and the sum is restricted to the equilibria with $f^a = f$. We note that $\bar{\phi}(f)$ depends on the specific realization of the connectivity matrix $\mathbf{w}$.

To determine the angle, $\omega^a(f)$, between ϕ^a and $\bar{\phi}(f)$, we evaluate their scalar product. This is given by

$$\phi^a \cdot \bar{\phi}(f) = \|\phi^a\| \|\bar{\phi}(f)\| \cos \omega^a(f) = \sum_i \phi_i^a \frac{1}{P_{\mathbf{w}}(f)} \sum_{b|f^b=f} \phi_i^b = \frac{1}{P_{\mathbf{w}}(f)} \sum_{b|f^b=f} \sum_i \phi_i^a \phi_i^b \quad (\text{S-99})$$

$$= \frac{N}{P_{\mathbf{w}}(f)} (Q(f) + (P_{\mathbf{w}}(f) - 1) q(f)) = N q(f); \quad (N \rightarrow \infty) \quad (\text{S-100})$$

The norm of ϕ^a is simply $\|\phi^a\| = \sqrt{NQ(f)}$; the norm of $\bar{\phi}(f)$ is

$$\|\bar{\phi}(f)\|^2 = \sum_i \left(\frac{1}{P_{\mathbf{w}}(f)} \sum_{a|f^a=f} \phi_i^a \right)^2 = \frac{1}{P_{\mathbf{w}}^2(f)} \sum_{f^a=f^b=f} \sum_i \phi_i^a \phi_i^b \quad (\text{S-101})$$

$$= \frac{N}{P_{\mathbf{w}}^2(f)} (P_{\mathbf{w}}(f) Q(f) + P_{\mathbf{w}}(f) (P_{\mathbf{w}}(f) - 1) q(f)) = N q(f); \quad (N \rightarrow \infty) \quad (\text{S-102})$$

and, putting all together, we find

$$\cos \omega^a(f) = \sqrt{\frac{q(f)}{Q(f)}} \equiv \cos \omega(f) \quad (\text{S-103})$$

which does not depend on a , nor on $\mathbf{w}$, because both $Q(f)$ and $q(f)$ are self-averaging. Therefore, all ϕ^a 's with $f^a = f$ form the same angle with $\bar{\phi}(f)$. This implies that the ϕ^a 's are contained in a spherical cap, whose pole is determined by the intersection, in the non-negative orthant, between the straight line through the origin defined by $\bar{\phi}(f)$, and the hypersphere of radius $D(f) = \sqrt{N Q(f)}$ centered at the origin; the polar angle is given by $\omega(f)$. In fact, in the limit $N \rightarrow \infty$, the volume of the cap is concentrated in the rim, and any vector between the origin and a point in the rim makes the same angle with $\bar{\phi}(f)$, that is, $\omega(f)$. Note that this does not implies that all such vectors are equilibria, but rather that all the equilibria ϕ^a (with $f^a = f$) are *inside* the rim.

For ‘random’ equilibria (as defined above), the polar angle, $\hat{\omega}(f)$, is

$$\cos \hat{\omega}(f) = \sqrt{\frac{1}{Q(f)}} < \cos \omega(f) \quad (\text{S-104})$$

and, hence, $\hat{\omega}(f) > \omega(f)$ for all f . Thus, actual equilibria can be enclosed in a volume (much) smaller than the volume needed to enclose ‘random’ equilibria. This is another manifestation of the fact that actual equilibria are correlated.

The above arguments suggest that we also should expect significant correlations between equilibria with different fractions of active neurons. These can be quantified by introducing the overlaps, $q^{ab}(f_1, f_2)$, between two equilibria, ϕ^a with $f^a = f_1$ and ϕ^b with $f^b = f_2$. These are defined by

$$q^{ab}(f_1, f_2) = q(f_1, f_2) \equiv \frac{1}{N} \sum_i \phi_i^a \phi_i^b \quad (\text{S-105})$$

where we have dropped the dependence on the pair of solutions, because we expect $q^{ab}(f_1, f_2)$ to be self-averaging. In fact, this can be seen more ‘rigorously’, and $q(f_1, f_2)$ can be computed analytically, by a relatively straightforward extension of the (replica-symmetric) theory presented in the previous section.

For ‘random’ equilibria, $q(f_1, f_2) = 1$. In fact, the probability that both ϕ_i^a and ϕ_i^b are different from zero is $f_1 f_2$, while $\langle \phi_i^a \rangle_A = 1/f_1$ and $\langle \phi_i^b \rangle_A = 1/f_2$. Thus, $q(f_1, f_2) > 1$ indicates significant pair-wise correlations among equilibria with two different fractions, f_1 and f_2 , of active neurons.

As discussed above, both the ϕ^a 's and the ϕ^b 's are inside spherical caps. The relative position of these spherical caps can be conveniently quantified by the cosine similarity between the corresponding polar vectors; this is given by

$$\cos \omega(f_1, f_2) \equiv \frac{q(f_1, f_2)}{\sqrt{q(f_1) q(f_2)}} \quad (\text{S-106})$$

where $\omega(f_1, f_2)$ is the angle between the center-of-mass of the ϕ^a 's with $f^a = f_1$, $\bar{\phi}(f_1)$, and the center-of-mass of the ϕ^b 's with $f^b = f_2$, $\bar{\phi}(f_2)$.

III. DYNAMICAL MEAN-FIELD THEORY

Here we summarize the dynamical mean-field theory (DMFT) for our model network. Our presentation is short and *ad rem*. Detailed derivations of the DMFT can be found in the literature; a very detailed and pedagogical one is [3].

We are concerned only with equilibrium, i.e., when network dynamics has reached a steady state. We define

$$\Delta(\tau) \equiv \frac{1}{N} \sum_{i=1}^N (x_i(t) - \mu)(x_i(t + \tau) - \mu) \quad (\text{S-107})$$

where $\mu \equiv N^{-1} \sum_{i=1}^N x_i(t)$, which is constant in time (at equilibrium), and

$$q(\tau) \equiv \frac{1}{N} \sum_{i=1}^N \phi_i(t) \phi_i(t + \tau) \quad (\text{S-108})$$

Clearly, $\Delta(\tau) = \Delta(-\tau)$ and $q(\tau) = q(-\tau)$. For later convenience, we introduce the following short-hand notations: $\Delta_0 \equiv \Delta(0)$, $\Delta_\infty \equiv \lim_{\tau \rightarrow \infty} \Delta(\tau)$ and, analogously, for q_0 and q_∞ (corresponding to Q_D and q_D in the main text). Obviously, $\Delta_0 \geq \Delta_\infty$ and $q_0 \geq q_\infty$.

Two self-consistency conditions can be immediately derived. From the balance condition, we obtain

$$1 = \int Dz \phi(\mu + z\sqrt{\Delta_0}) \quad (\text{S-109})$$

which expresses the fact that the average firing rate in the network is constant in time and determined by the balance condition. In the limit $\tau \rightarrow \infty$, $x_i(t)$ and $x_i(t + \tau)$ become uncorrelated. Thus,

$$\Delta_\infty = \sigma_w^2 \int Dz \left(\int Dt \phi(\mu + t\sqrt{\Delta_0 - \Delta_\infty} + z\sqrt{\Delta_\infty}) \right)^2 \quad (\text{S-110})$$

We note that when the network dynamics is in an equilibrium, i.e., $\Delta_0 = \Delta_\infty$ and, hence, $q_0 = q_\infty$, Eqs. (S-109) and (S-110) are the same as the ‘cavity’ solution. In the general case, however, we need a third self-consistent equation that has no (obvious) counterpart in the ‘static’ theory.

The third self-consistent condition can be obtained as follows: $\Delta(\tau)$ and $q(\tau)$ are related by the following equation

$$\ddot{\Delta} = \Delta - \sigma_w^2 q = -\frac{\partial V}{\partial \Delta} \quad (\text{S-111})$$

where the double dot denotes the second derivative with respect to time (i.e., τ) and $V(\Delta)$ is a ‘potential’, which can always be defined because the dynamics is 1-D. Clearly, the solutions of Eq. (S-111) conserve the

total ‘energy’, i.e.,

$$\frac{1}{2}\dot{\Delta}^2 + V(\Delta) = \text{constant} \quad (\text{S-112})$$

along all the orbits of Eq. (S-111). In the limit $\tau \rightarrow \infty$, $\dot{\Delta}(\tau) = 0$. Also, $\Delta(\tau)$ is symmetric around 0 and reaches its maximum value there, hence $\dot{\Delta}(0) = 0$. Using this information into Eq. (S-112), we obtain

$$V(\Delta_0) = V(\Delta_\infty) \quad (\text{S-113})$$

because the ‘kinetic energy’ is 0 both at $\tau = 0$ and at $\tau \rightarrow \infty$. Eq. (S-113) provides the required third self-consistent condition. The potential V can be explicitly computed (see [3]); in our case, it reads

$$V(\Delta) = -\frac{\Delta^2}{2} + \sigma_w^2 \int Dz \left(\int Dt \Phi \left(\mu + t\sqrt{\Delta_0 - \Delta} + z\sqrt{\Delta} \right) \right)^2 \quad (\text{S-114})$$

where $\Phi(x)$ is the anti-derivative of $\phi(x)$. In our case,

$$\Phi(x) = \frac{x^2}{2} \Theta(x) \quad (\text{S-115})$$

IV. NUMERICAL SIMULATIONS

A. Newton’s method with backtracking

We use Newton’s method with a backtracking line search to find equilibria of the dynamics. At each iteration, the update is computed as

$$\Delta \mathbf{x} = -\mathbf{J}^{-1}(\mathbf{x}) \left(\sqrt{N} - \mathbf{x} + \frac{1}{\sqrt{N}} \mathbf{w} \phi \right), \quad \mathbf{x} \rightarrow \mathbf{x} + \alpha \Delta \mathbf{x}, \quad (\text{S-116})$$

where $\mathbf{J}(\mathbf{x})$ is the Jacobian of the dynamical equation, and $\phi = \phi(\mathbf{x})$ is the nonlinear activation. The update is accepted if it decreases the objective function, defined as the norm of the velocity, $f(\mathbf{x}) = |\sqrt{N} - \mathbf{x} + \frac{1}{\sqrt{N}} \mathbf{w} \phi|$. Otherwise, the step size α (initialized as 1) is reduced by a factor of 0.5 until the condition below is satisfied

$$f(\mathbf{x} + \alpha \Delta \mathbf{x}) \leq f(\mathbf{x}) + c(f(\mathbf{x} + \Delta \mathbf{x}) - f(\mathbf{x})). \quad (\text{S-117})$$

We set $c = 0.001$ as the backtracking parameter and set the maximum number of iterations to 1000. The algorithm stops when $f(\mathbf{x})$ is smaller than a predefined threshold, $f(\mathbf{x}) < 10^{-3}$, or when the maximum number of iterations is reached. For each network realization, we run Newton’s method with 10^6 - 10^7 different initial conditions, independently and uniformly sampled within an N -dimensional hypercube with large enough side length, i.e., $L = \sqrt{N}$. This choice ensures that the entire phase space is explored uniformly, without bias toward any particular region.

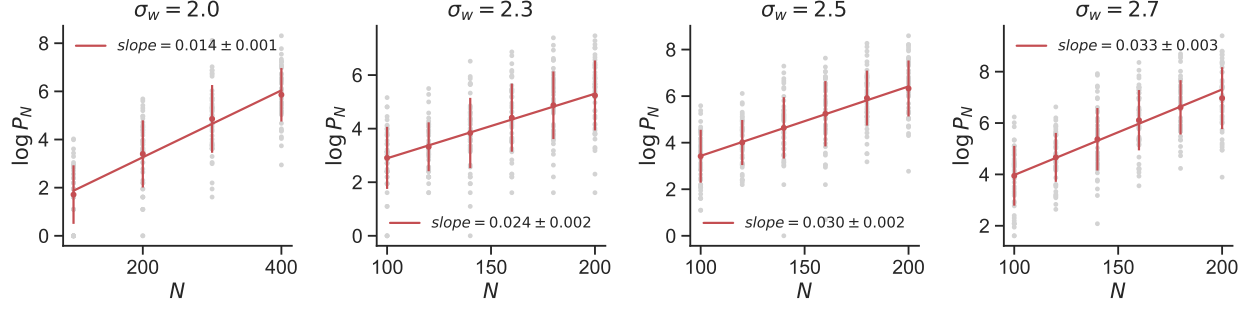


FIG. S1. Numerical estimation of Σ_q^* . Grey dots: $\log P$ for $M = 50$ sample networks. Error bars: mean and SD across sample networks.

A.1. Determining the number of initial conditions

To reliably identify all distinct equilibria, we do not fix the number of initial conditions *a priori*. Instead, we choose it adaptively to ensure that each unique equilibrium (see below for the definition of uniqueness) is, on average, discovered at least m times from independent searches. Larger values of m improve the robustness of the complexity estimate, but also increase computational cost. In our numerical search, we set $m = 30$ as a practical compromise between accuracy and efficiency.

A.2. Identifying true and unique equilibria

To compare the numerical results with theory, it is crucial to accurately count the number of true and unique equilibria identified by the algorithm. In practice, it is challenging to distinguish a true equilibrium from a slow point (a location where the velocity is small but nonzero), and to decide whether two nearby equilibria should be treated as identical within finite machine precision. This problem can be addressed by using threshold-linear activation units. In this case, each equilibrium is uniquely characterized by a binary configuration vector, $\mathbf{s} = \Theta(\mathbf{x})$, and the corresponding criterion can be written as

$$(2s_i - 1)h_i + \frac{1}{\sqrt{N}} \sum_{j=1}^N w_{ij}s_j h_i = \sqrt{N}, \quad i = 1 \dots N, \quad (\text{S-118})$$

where $s_i = \Theta(x_i)$ and $h_i = |x_i|$. $\mathbf{x}$ is classified as a true equilibrium if the solution exists and all components lie in the positive orthant, i.e., $h_i \geq 0$ for all $i = 1 \dots N$.

A.3. Numerical estimation of complexity

The algorithm allows us to compare theory with estimates of Σ_q^* obtained by counting the number of equilibria in sample networks. For each pair (σ_w, N) , we generate $M = 50$ network realizations and search for their equilibria using the algorithm described above. For fixed σ_w , we regress $\log P$ against N (with M samples of $\log P$ for each N) to obtain an estimate of Σ_q^* at that σ_w . The results of this procedure are shown in Fig. S1.

B. Numerical estimation of order parameters

Let us consider a given sample network and let ϕ^a ($a = 1, \dots, P$) be its equilibria. We define $S(n)$, with $n = 0, \dots, N$, the set of the indexes of the equilibria which have exactly n neurons active, i.e.,

$$S(n) \equiv \left\{ a : \sum_i \Theta(x_i^a) = n, 1 \leq a \leq P \right\} \quad (\text{S-119})$$

Note that there will be n for which $S(n) = \emptyset$, i.e., $S(n)$ is the empty set. The upcoming analysis is restricted to values of n for which $S(n)$ contains at least one element. Clearly,

$$\bigcup_{n=0}^N S(n) = \{1, \dots, P\}; \quad S(n_1) \cap S(n_2) = \emptyset \quad \text{if } n_1 \neq n_2 \quad (\text{S-120})$$

We denote $P(n)$ the number of elements in $S(n)$, i.e., its cardinality, $P(n) \equiv |S(n)|$. $P(n) = 0$ if $S(n)$ is empty. Clearly,

$$\sum_{n=0}^N P(n) = P \quad (\text{S-121})$$

We consider values of n for which $S(n)$ is not empty. For these n , we compute

$$Q^a(n) = \frac{1}{N} \sum_i (\phi_i^a)^2; \quad a \in S(n) \quad (\text{S-122})$$

From these $P(n)$ quantities, we compute $\hat{Q}(n)$ as

$$\hat{Q}(n) = \frac{1}{P(n)} \sum_{a \in S(n)} Q^a(n) \quad (\text{S-123})$$

Thus, $\hat{Q}(n)$ is a tabulated estimate of $Q(f)$ ($n = fN$) in a *given* sample network.

We can similarly obtain a tabulated estimate of $q(f)$, $\hat{q}(n)$, in a given sample network. For this, we compute

$$q^{ab}(n) = \frac{1}{N} \sum_i \phi_i^a \phi_i^b; \quad a, b \in S(n), \quad a < b \quad (\text{S-124})$$

There are $\frac{1}{2}P(n)(P(n) - 1)$ such quantities, from which we compute

$$\hat{q}(n) = \frac{2}{P(n)(P(n) - 1)} \sum_{\substack{a, b \in S(n) \\ a < b}} q^{ab}(n) \quad (\text{S-125})$$

One can also obtain a tabulated estimate of the overlap between two equilibria with different f (similar to

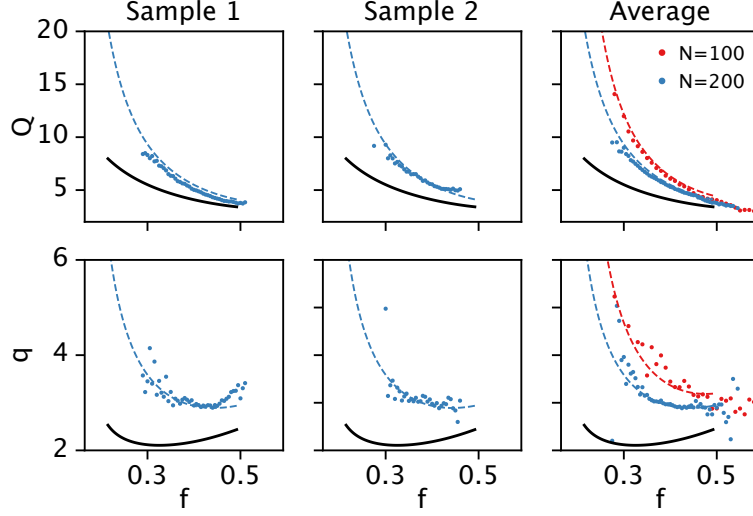


FIG. S2. Estimated order parameters $Q(f)$ and $q(f)$ of two randomly selected sample networks (first two columns) and averaged over 50 sample networks (last column) for $\sigma_w = 2.5$. Full black lines: asymptotic theory; Dashed lines: finite- N theory; Dots: numerical simulations.

our analysis of the centroids). For this one computes

$$q^{ab}(n_1, n_2) = \frac{1}{N} \sum_i \phi_i^a \phi_i^b; \quad a \in S(n_1), \quad b \in S(n_2) \quad (\text{S-126})$$

There are $P(n_1)P(n_2)$ such quantities, from which one computes

$$\hat{q}(n_1, n_2) = \frac{1}{P(n_1)P(n_2)} \sum_{\substack{a \in S(n_1) \\ b \in S(n_2)}} q^{ab}(n_1, n_2) \quad (\text{S-127})$$

It would be interesting to see how these tabulated estimates fluctuate across sample networks. The simplest is to look at $\hat{Q}_m(n)$ and $\hat{q}_m(n)$, where $m = 1, \dots, M$ ($M = 50$) now denotes the sample network from which the estimates are obtained. To start, one can just look at the case $N = 200$.

For instance, for a given σ_w (say, $\sigma_w = 2.5$), one can plot $\hat{Q}_m(n)$ against n for all m on the same figure (each m with a different color). Similarly, in another figure and for the same σ_w , one can plot $\hat{q}_m(n)$ against n for all m .

Finally, one can plot (in a different figure) the same quantities averaged across sample networks, i.e.,

$$\bar{Q}(n) = \frac{1}{M} \sum_m \hat{Q}_m(n); \quad \bar{q}(n) = \frac{1}{M} \sum_m \hat{q}_m(n) \quad (\text{S-128})$$

$\overline{Q}(n)$ and $\overline{q}(n)$ should be a fairly good estimate of $Q(f)$ and $q(f)$, respectively.

-
- [1] Kevin Berlemont and Gianluigi Mongillo. Glassy phase in dynamically-balanced neuronal networks. *bioRxiv*, 2022.
 - [2] Yashar Ahmadian, Francesco Fumarola, and Kenneth D Miller. Properties of networks with partially structured and partially random connectivity. *Physical Review E*, 91(1):012820, 2015.
 - [3] A Crisanti and H Sompolinsky. Path integral approach to random neural networks. *Physical Review E*, 98(6):062120, 2018.