# Assessing the Similarity of Real Matrices with Arbitrary Shape

Jasper Albers [1,2,*,†] Anno C. Kurth [1,2,*,‡] Robin Gutzen [1] Aitor Morales-Gregorio [1,3] Michael Denker [1] Sonja Grün [1,4,5] Sacha J. van Albada [1,6] and Markus Diesmann [1,4,7,8]

[1] *Institute for Advanced Simulation (IAS-6), Jülich Research Centre, 52425 Jülich, Germany*

[2] *RWTH Aachen University, 52062 Aachen, Germany*

[3] *Faculty of Mathematics and Physics, Charles University, 180 00 Prague, Czech Republic*

[4] *JARA-Institut Brain Structure-Function Relationships (INM-10), Jülich Research Centre, 52425 Jülich, Germany*

[5] *Theoretical Systems Neurobiology, RWTH Aachen University, 52056 Aachen, Germany*

[6] *Institute of Zoology, University of Cologne, 50674 Cologne, Germany*

[7] *Department of Psychiatry, Psychotherapy and Psychosomatics, School of Medicine, RWTH Aachen University, 52074 Aachen, Germany*

[8] *Department of Physics, Faculty 1, RWTH Aachen University, 52074 Aachen, Germany*

(Received 10 April 2024; revised 6 May 2024; accepted 10 March 2025; published 5 May 2025)

Assessing the similarity of matrices is valuable for analyzing the extent to which data sets exhibit common features in tasks such as data clustering, dimensionality reduction, pattern recognition, group comparison, and graph analysis. Methods proposed for comparing vectors, such as cosine similarity, can be readily generalized to matrices. However, this approach usually neglects the inherent two-dimensional structure of matrices. Here we propose *singular angle similarity* (SAS), a measure for evaluating the structural similarity between two arbitrary, real matrices of the same shape based on singular value decomposition. After introducing the measure, we compare SAS with standard measures for matrix comparison and show that only SAS captures the two-dimensional structure of matrices. Further, we characterize the behavior of SAS in the presence of noise, as a function of matrix dimensionality, and when singular values are degenerate. Finally, we apply SAS to two use cases: square nonsymmetric matrices of probabilistic network connectivity, and nonsquare matrices representing neural brain activity. For synthetic data of network connectivity, SAS matches intuitive expectations and allows for a robust assessment of similarities and differences. For experimental data of brain activity, SAS captures differences in the structure of high-dimensional responses to different stimuli. We conclude that SAS is a suitable measure for quantifying the shared structure of matrices with arbitrary shape.

DOI: 10.1103/PRXLife.3.023005

## I. INTRODUCTION

Social groups, transportation systems, chemical reactions, brains—many complex systems are governed and commonly characterized by the pairwise interactions of their constituent elements. Typically, these interactions are described by matrices that can represent covariance structures, spatiotemporal dependencies, or the connections and interactions in a network, forming the foundation for the mathematical treatment of such complex systems [1]. Common examples include genetic variance [2], ecological food chains [3], or stock markets [4]. Additionally, many other types of structured data can be represented in matrix form, ranging from test scores for groups of subjects to parallel time series data. Quantifying the similarity between such matrices is important for distinguishing features of the underlying systems. Examples include the representations of stimuli in artificial and biological neural networks. The extent to which such representations are similar helps to address questions regarding convergent learning, i.e., whether solutions found by neural networks are universal or specific. Representational similarity has been investigated across different neural network architectures [5,6], different training data [6], and different initializations of the parameters [7]. In the context of biological neural networks, differences in the representation within brain areas [8], between species [9] as well as between brains and artificial neural networks [10–12] have been studied. See [13] for a review of the topic.

Classical measures of the similarity between two matrices $A$ and $B$ are often based on the Frobenius scalar product $\langle A, B\rangle_F = \mathrm{tr}(AB^T)$, leading to the Frobenius norm $||A - B||_F^2 = \langle A - B, A - B\rangle_F$, or the cosine similarity $\langle A, B\rangle_F$ where $||A||_F = ||B||_F = 1$ [14]. However, the Frobenius norm and cosine similarity take into account only the numerical values of corresponding entries of the matrices and ignore their two-dimensional structure. For non-negative matrices, information theory-inspired approaches have been suggested [15,16]. These are conceptually similar in the sense that also here the relative position of matrix entries is ignored.

*These authors contributed equally to this work.
†Contact author: j.albers@fz-juelich.de
‡Contact author: a.kurth@fz-juelich.de

Measures partially overcoming the defect of ignoring the two-dimensional structure of matrices include the centered kernel alignment [6,17] (CKA) and the normalized Bures similarity [18]. For linear kernels, CKA reads as $||AB^T||_F^2/||AA^T||_F||BB^T||_F$. Other popular choices for CKA kernels include radial basis functions. The normalized Bures similarity is identical to CKA for linear kernels with the exception that the Frobenius norm is replaced by the nuclear norm. Kernel-based approaches [19–22] are also used for assessing the similarity between graphs (represented by their adjacency matrices; see Kriege *et al.* [23] for a recent survey).

Other authors propose comparisons based on correlation coefficients or the coefficient of determination as a means to assess similarity [8,9]. This approach is not limited to standard Pearson correlation coefficients but instead can also employ regularized regression analyses such as Ridge or LASSO regression.

More involved approaches reformulate ideas borrowed from canonical correlation analysis (CCA) [24] in the context of the comparison of two matrices [25]. They rely on the canonical angles between the subspaces spanned by the columns of the matrices. Here, problems may arise if the embedding space (the number of rows) is of similar size as the number of columns and, at the same time, the subspaces spanned by these columns are high-dimensional (relative to the dimension of the embedding space). In this case, the canonical angles cannot meaningfully distinguish the subspaces, limiting the applicability of these approaches. Alternatively, similarity measures based on Procrustes analysis have been proposed [26]. Assuming that matrices $A$ and $B$ are centered, the Procrustes problem asks for the minimum of $||OA - B||_F$ where $O$ is an orthogonal matrix [27]. This minimum can be interpreted as a similarity and has a direct geometric meaning when viewing the matrix elements as coordinates of points in space: the Procrustes similarity assesses to which extent two objects, represented by the matrices $A$ and $B$, overlap after reflecting and rotating them so that the overlap is maximized. Other formulations consider a measure like the cosine similarity after having applied the optimal orthogonal transformation [28]. See Ding *et al.* [29] or Cloos *et al.* [30] for a comparison of some of the mentioned similarity measures. Finally, for symmetric positive-definite (SPD) matrices (e.g., covariance matrices), similarity measures based on geodesic distances induced by a natural Riemannian metric on the space of SPD matrices have been studied [31,32]. The induced distance is invariant under the action of the general linear group on the SPD matrices and thus especially under joint diagonalization.

Previous work addresses the comparison of symmetric matrices using eigenangles—the angles between eigenvectors of the compared matrices [33]. Small eigenangles indicate a good alignment of the respective eigenspaces, enabling the definition of a similarity score. The authors employ an analytical description of the similarity score based on random matrix theory to devise a statistical test for the comparison of such matrices. Asymmetric matrices usually have complex eigenvectors and eigenvalues, making their ordering ambiguous. Thus, an extension of the eigenangle test to asymmetric matrices is not straightforward. Additionally, the approach implicitly assumes that all eigenvalues are nondegenerate. Finally, the eigenangle test is by definition not applicable to nonsquare matrices. In this study, we overcome these limitations by proposing a refined matrix similarity measure that naturally extends to the comparison of any two real matrices with identical shapes. Using singular value decomposition (SVD) instead of eigendecomposition, we derive SAS from the respective left and right singular vectors and singular values. In this way, we vastly generalize the approach introduced in [33] and enable a multitude of applications not possible previously.

In Sec. II we formally define SAS and derive basic properties. Further, three types of matrices are introduced that we use in the following for the evaluation of SAS and comparison to other similarity measures: random matrices with continuously distributed entries (Sec. II B), adjacency matrices of random graphs (Sec. II C), and massively parallel neural activity recordings (Sec. II D). Finally, we detail how degeneracy is gradually introduced to the singular value spectrum of matrices (Sec. II E).

We start our analysis by providing a geometric interpretation of SAS (Sec. III A). Next, we compare SAS with standard similarity measures for matrices: on the basis of generic random matrices we show that only SAS captures certain salient two-dimensional correlation structures (Sec. III B). This comparison follows a templated structure where we first calculate the self-similarity distribution of different samples of one random matrix class assessed by SAS. This distribution is then compared with the cross-similarity distribution of SAS scores from sampled random matrices of the given class with other random matrix classes. Third, we characterize the behavior of SAS under changes in dimensionality, perturbation of entries, and degeneracy of singular values (Sec. III C). Fourth, we evaluate the similarity across instances of six probabilistic graph models that are commonly used to describe network architecture (Sec. III D). With this use case, we demonstrate that SAS is able to differentiate between the connectivity in network graphs by means of their adjacency matrices. Finally, we apply SAS on experimental data, evaluating the similarity of brain activity in the visual cortex of macaques in response to four different visual stimuli (Sec. III E). This application shows that SAS can identify underlying features in the presence of realistic noise.

In conclusion (Sec. IV), we show that SAS is a well-behaved measure for structural similarity in matrices that is applicable in different scientific domains. It highlights shared variability between matrices and allows for a distinction of models or processes underlying their generation.

## II. METHODS

### A. Singular angle similarity

To assess the similarity of two arbitrary, real, $m \times n$ matrices $M_a, M_b$, we devise a measure based on singular value decomposition (SVD) [34]. Without loss of generality, we assume $m \leqslant n$. SVD guarantees the existence of orthogonal matrices $U_i \in \mathbb{R}^{m\times m}$ and $V_i \in \mathbb{R}^{n\times n}$, and diagonal matrices $\Sigma_i = \mathrm{diag}(\sigma_i^1, \ldots, \sigma_i^m) \in \mathbb{R}^{m\times n}$ where $\sigma_i^j \geqslant \sigma_i^l \geqslant 0$ for $l > j \geqslant 1$ such that

$$M_i = U_i \Sigma_i V_i^T . \tag{1}$$

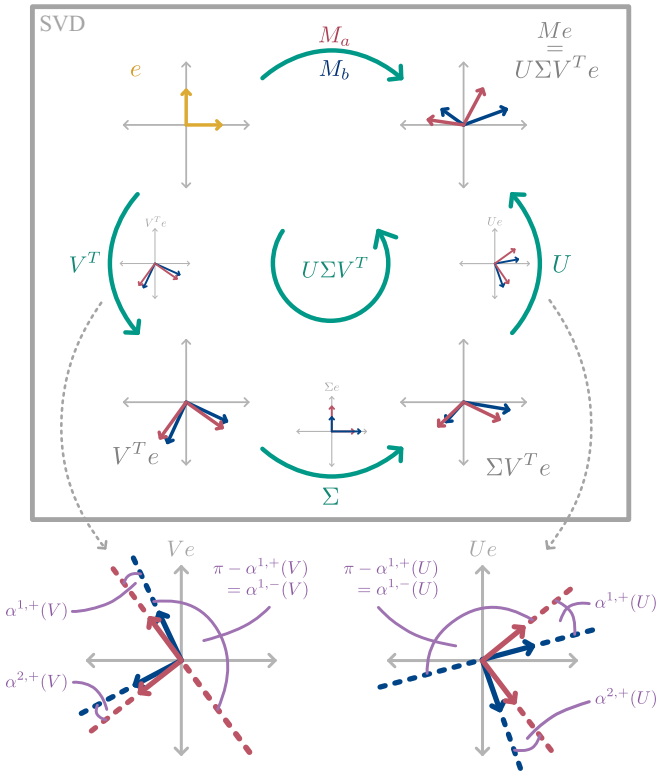


FIG. 1. Singular value decomposition and singular angles. Schematic representation of the transformations applied on the basis vectors $e$ (yellow) by the components of SVD for two $2 \times 2$ matrices $M_a$ (red), and $M_b$ (blue). Small graphs next to the green arrows illustrate the isolated action of the corresponding transformations $V^T$, $\Sigma$, and $U$ on $e$. Below the singular angles of the first $[\alpha^{j,+}(V)$ and $\alpha^{j,+}(U)]$ and second $[\alpha^{j,-}(V)$ and $\alpha^{j,-}(U)]$ kind are shown as the angles between the column vectors of $V_a$ and $V_b$, and $U_a$ and $U_b$, respectively, as defined in Eq. (3).

Here $i \in \{a, b\}$. SVD is schematically presented for a $2 \times 2$ case in Fig. 1. The columns of $V_i$, denoted by $v_i^1, \ldots, v_i^n$, are the right singular vectors, and the columns of $U_i$, denoted by $u_i^1, \ldots, u_i^m$, are the left singular vectors. With this, the SVD can also be written in the form

$$M_i = \sum_{j=1}^{m} \sigma_i^j u_i^j \otimes v_i^j. \tag{2}$$

Here $\otimes$ denotes the outer product of two vectors. Thus, under the action of $M_i$ the vectors $v_i^j$ are transformed into the vectors $\sigma_i^j u_i^j$. The singular values $\sigma_i^1, \ldots, \sigma_i^m$ are unique—they are the square root of the eigenvalues of $M_i M_i^T \in \mathbb{R}^{m \times m}$. We note that if both $M_i$ are symmetric positive-definite matrices (e.g., covariance matrices), then $U_i = V_i$, and the singular values become the eigenvalues.

For simplicity of the derivations, we at first assume that there are no degenerate singular values except zero. The overwhelming majority of higher-dimensional matrices encountered in practice indeed satisfy this assumption.

Left and right singular vectors that correspond to nondegenerate singular values are uniquely determined up to a joint multiplication by $-1$. Thus, the vector pairs $(u_i^j, v_i^j)$ and $(-u_i^j, -v_i^j)$ both are equally valid singular vectors to a nondegenerate singular value $\sigma_i^j \geqslant 0$.

If $\min\{\sigma_a^j, \sigma_b^j\} \neq 0$ we define the left singular angle $\alpha^{j,+}(U)$ and the right singular angle $\alpha^{j,+}(V)$ of the first kind as

$$\begin{aligned} \alpha^{j,+}(U) &= \sphericalangle\left(u_a^j, u_b^j\right) = \arccos\left(\left|u_a^j, u_b^j\right|\right), \\ \alpha^{j,+}(V) &= \sphericalangle\left(v_a^j, v_b^j\right) = \arccos\left(\left|v_a^j, v_b^j\right|\right). \end{aligned} \tag{3}$$

Due to the ambiguity in vector pairs of left and right singular vectors we additionally define left singular angles of the second kind as $\alpha^{j,-}(U) = \sphericalangle(-u_a^j, u_b^j) = \sphericalangle(u_a^j, -u_b^j)$ and *mutatis mutandis* for right singular angles of the second kind $\alpha^{j,-}(V)$. The singular angles of the first and second kinds are visualized in Fig. 1. There we have

$$\alpha^{j,+}(U) + \alpha^{j,-}(U) = \pi = \alpha^{j,+}(V) + \alpha^{j,-}(V). \tag{4}$$

Due to the ambiguity in the sign, one has to consider either $(\alpha^{j,+}(U), \alpha^{j,+}(V))$ or $(\alpha^{j,-}(U), \alpha^{j,-}(V))$ together. We define the *singular angle* as the smaller average of the two choices

$$\begin{aligned} \alpha^j &= \min\left\{\frac{\alpha^{j,+}(U) + \alpha^{j,+}(V)}{2}, \frac{\alpha^{j,-}(U) + \alpha^{j,-}(V)}{2}\right\} \\ &= \min\{\bar{\alpha}^j, \pi - \bar{\alpha}^j\}, \end{aligned} \tag{5}$$

where $\bar{\alpha}^j = [\alpha^{j,+}(U) + \alpha^{j,+}(V)]/2$. Using the angular similarity

$$\Delta^j = 1 - \frac{\alpha^j}{\pi/2} \in [0, 1] \tag{6}$$

and defining the *singular value score* as $w^j = w(\sigma_a^j, \sigma_b^j)$ where $w(x, y) \geqslant 0$ denotes a weight function, we calculate SAS as the weighted average of the angular similarities:

$$\mathrm{SAS} = \frac{\sum_j^k w^j \Delta^j}{\sum_j^k w^j} \in [0, 1]. \tag{7}$$

Here $k$ is the largest natural number less than or equal to $m$ such that $\min\{\sigma_a^k, \sigma_b^k\} \neq 0$. In the following, we choose

$$w(x, y) = (x + y)/2. \tag{8}$$

Other possible choices include $w(x, y) = \sqrt{(x^2 + y^2)/2}$ (cf. [33]) and $w(x, y) = \sqrt{xy}$. One can substitute other vector-based similarity measures for the angular similarity defined in Eq. (6). For example, substituting cosine similarity yields $\Delta^j = \cos(\alpha^j)$.

According to our definition of SAS in Eq. (7), singular angles stemming from singular vectors of which at least one has a corresponding singular value of zero do not contribute to SAS. By definition, SAS can be computed only for matrices of the same shape.

*a. Degenerate case.* If two or more nonzero singular values are identical, the described approach cannot be readily applied since there is no canonical choice for pairing singular vectors: left and right singular vectors $u_i^k, \ldots, u_i^{k+l}$ and $v_i^k, \ldots, v_i^{k+l}$ corresponding to the degenerate singular value $\sigma = \sigma_i^k = \cdots = \sigma_i^{k+l}$ are only unique up to an orthogonal transformation acting on the subspaces spanned by the singular vectors. This is the higher-dimensional generalization of

the ambiguity regarding the factor $-1$ for the one-dimensional case described above. Writing the corresponding vectors as columns of matrices $U_i^{[k:k+l]} \in \mathbb{R}^{n\times l}$ and $V_i^{[k:k+l]} \in \mathbb{R}^{m\times l}$, this means that the columns of the matrices

$$U_i^{[k:k+l]}O \quad \text{and} \quad V_i^{[k:k+l]}O \tag{9}$$

are valid left and right singular vectors. Here $O \in \mathbb{R}^{l\times l}$ is an arbitrary orthogonal matrix. Note that in order to maintain consistency between the singular vectors, the same orthogonal transformation $O$ must be applied to both, $U_i^{[k:k+l]}$ and $V_i^{[k:k+l]}$. Thus, left and right singular vectors can be given only up to this ambiguity, and consequently singular angles according to Eq. (3) are not well defined.

This problem can be mitigated by using the *canonical angles* between the subspaces spanned by the left and right singular vectors of the matrices $M_i$ [35]. If the two subspaces are given in terms of orthonormal bases $B_i$ (written again as columns of matrices), the canonical angles are the angles between corresponding column vectors of $B_aO_a$ and $B_bO_b$, where $O_a, O_b$ are suitable orthogonal transformations [36] (see Sec. S1.1 [37]). This can be interpreted as optimally aligning the two orthonormal coordinate systems while keeping the subspaces invariant.

We define singular angles for degenerate singular values building on that interpretation: assuming without loss of generality $\sigma = \sigma_a^k = \cdots = \sigma_a^{k+l}$, we define $U$-aligned left and right singular angles for $k \leqslant j \leqslant k+l$ as

$$\begin{aligned}\alpha^{j,U}(U) &= \sphericalangle\big(U_a^{[k:k+l]}O_a^U e_j, U_b^{[k:k+l]}O_b^U e_j\big),\\ \alpha^{j,U}(V) &= \sphericalangle\big(V_a^{[k:k+l]}O_a^U e_j, V_b^{[k:k+l]}O_b^U e_j\big).\end{aligned} \tag{10}$$

Here $e_j$ is the $j$th standard normal basis vector, and the orthogonal transformations $O_a^U, O_b^U$ are chosen such that $\alpha^{j,U}(U)$ are the canonical angles. The $U$-aligned singular angles are then given by the mean of the $U$-aligned left and right singular angles:

$$\alpha^{j,U} = \frac{\alpha^{j,U}(U) + \alpha^{j,U}(V)}{2}. \tag{11}$$

The corresponding angles in the $V$-aligned case are defined *mutatis mutandis*. The singular angles are then given by either the $U$- or $V$-aligned singular angles, depending on which have the smaller sum:

$$\alpha^j = \alpha^{j,X} \text{ where } X = \begin{cases} U & \text{if } \sum_j \alpha^{j,U} < \sum_j \alpha^{j,V} \\ V & \text{if } \sum_j \alpha^{j,V} \leqslant \sum_j \alpha^{j,U}. \end{cases} \tag{12}$$

In this way, the singular angles for singular vectors corresponding to degenerate singular values directly generalize the singular angles in the nondegenerate case. For their contribution to SAS, we define $w^j = w(\sigma_a^j, \sigma_b^j)$ for $k \leqslant j \leqslant k+l$, i.e., the singular values are combined in their matching order. A complication may arise when $\sigma_a^k = \cdots = \sigma_a^{k+l}$ and $\sigma_b^{k-m} = \cdots = \sigma_b^{k+n}$ for some $m$ and $n$. Here the degenerate subspaces of the two matrices are partially overlapping. We treat this case by applying the above described method to the subspaces given by $U_i^{[k-m:k+\max\{n,l\}]}, V_i^{[k-m:k+\max\{n,l\}]}$.

Additionally, if two singular values are close so that small perturbations can lead to a change in their order, the pairing of singular vectors will change, potentially leading to different SAS values. This can be avoided by rounding the singular values to a precision determined for the matrices at hand. Thereby, degeneracy is introduced which can be treated as described above.

*b. Large matrices.* For applications, it might be impractical to compute the full SVD of the matrices $M_a, M_b$ if $n, m$ are large. In this case, SAS can still be used when replacing the full SVD with the truncated SVD [38]: instead of decomposing the matrices exactly according to Eq. (1), one seeks to find a low-rank approximation:

$$M_i \approx {}^tU_i^t \Sigma_i^t V_i^T. \tag{13}$$

Here ${}^tU_i \in \mathbb{R}^{m\times k}$ and ${}^tV_i \in \mathbb{R}^{n\times k}$ have orthogonal columns and ${}^t\Sigma_i \in \mathbb{R}^{k\times k}$ is a diagonal matrix with non-negative entries. The singular values and corresponding singular vectors of the truncated SVD are identical to the $k$ largest singular values of the full SVD and their corresponding singular vectors. The order of the approximation, $k$, depends on the concrete application. Note that the Eckart-Young-Mirsky theorem asserts that this low-rank approximation is minimal with respect to the Frobenius norm given the order $k$ [39]. The decomposition in Eq. (13) can straightforwardly be used to compute SAS. The resulting similarity between matrices is still informative if the singular values decrease in magnitude sufficiently quickly. The validity of this assumption depends on the matrices at hand. In practice, oftentimes only few singular values, compared to the dimension of the matrices, are of large magnitude.

*c. Matrices of different shapes.* By definition, SAS can be applied only to matrices of identical shape. This poses limitations for situations in which matrices of different shapes occur naturally. Examples include data matrices from neuroscientific experiments involving behavioral paradigms where the time needed by a test animal for the completion of a task varies between trials and subjects. While such situations can in principle be addressed with various forms of "time warping" (see, e.g., [40] for the domain of neuroscience), similar transformations leading to satisfactory results may not exist in all contexts. To nonetheless assess the similarity, SAS can be extended by zero padding the singular vectors with smaller dimensionality so that the scalar products in Eq. (3) remain well defined. Thereby, dimensions present only in the vectors with larger dimensionality are neglected.

### B. Random matrices

To compare SAS to standard measures of matrix similarity, we define the following classes of random matrices with shape $N \times N$ where each entry is drawn from a continuous probability distribution. Such matrix ensemble are of great theoretical interest and of widespread use in various domains of science [41,42]. Numerical values for the corresponding model parameters are summarized in Table I.

*a. Uncorrelated normal matrix (UC).* For random matrices of this class, each entry is drawn independently from a normal distribution with the same mean $\mu$ and variance $\sigma^2$. Collections of these matrices are also referred to as a real Ginibre ensemble [43].

*b. Cross-correlated normal matrix (CC).* We first independently sample $N$ random vectors from an $N$-dimensional normal distribution with mean $\mu$ and covariance matrix $C$

TABLE I. Parameters of random matrices: uncorrelated normal matrix (UC), cross-correlated normal matrix (CC), cross-correlated block matrix (CB), and shuffled, cross-correlated block matrix (SB).

| Parameter | Model(s) | Meaning | Value |
|---|---|---|---|
| $N$ | All | Dimensionality | 300 |
| $\mu$ | All | Mean of distribution | 0 |
| $\sigma^2$ | UC | Variance of distribution | $1/N$ |
| $a$ | CC | Peak covariance | 10 |
| $b$ | CC | Inverse characteristic length | 100 |
| $b^{\text{lower}}$ | CB | Block index | 10 |
| $b^{\text{upper}}$ | CB | Block index | 90 |

TABLE II. Parameters of random graphs: Erdős-Rényi (ER), directed configuration model (DCM), one cluster (OC), two clusters (TC), Watts-Strogatz (WS), and Barabási-Albert (BA).

| Param. | Model(s) | Meaning | Value |
|---|---|---|---|
| $N_n$ | All | Number of nodes | 300 |
| $N_e$ | All | Number of possible connections | 90 000 |
| $N_c$ | All | Mean number of connections | 9000 |
| $r$ | OC | Relative increase of $p$ in cluster | 10 |
| $b^{\text{lower}}$ | OC | Cluster index | 50 |
| $b^{\text{upper}}$ | OC | Cluster index | 100 |
| $b^{\text{mid}}$ | TC | Index between clusters | 90 |
| $p_{\text{WS}}$ | WS | Reconnection probability | 0.3 |

where

$$C_{ij} = \frac{a}{N} \exp\left(-b\frac{|i-j|}{N}\right).$$

Thus, the entries of the correlation matrix decay exponentially with distance from the diagonal. The sampled vectors are the columns of an $N \times N$ matrix $K^1$. We repeat the process to obtain an independent matrix $K^2$ to finally define $K = (K^1 + K^{2T})/2$. Since each entry of the $K^i$ is normally distributed, so is each entry of their sum $K$, and the covariance between entries $K_{kl}$ and $K_{mn}$ is $(C_{kl} + C_{mn})/4$. Normalization by $N$ in the argument of the exponential ensures that the strength of the correlation scales with the size of the matrix.

*c. Cross-correlated block matrix (CB).* Again, uncorrelated normal (UC) matrices $B$ are sampled. Then the entries $B_{kl}$ where $b^{\text{lower}} \leqslant k \leqslant b^{\text{upper}}$ and $b^{\text{lower}} \leqslant l \leqslant b^{\text{upper}}$ are replaced by a correlated normal structure as defined above, forming a block on the diagonal.

*d. Shuffled, cross-correlated block matrix (SB).* Matrices are sampled according to CB. Then, the rows are permuted randomly while the columns remain untouched.

*e. Doubly shuffled, cross-correlated block matrix (DB).* Matrices are sampled according to CB. Then the rows and columns are permuted randomly.

### C. Random graphs

We compare the adjacency matrices of six well-known network models with SAS. For all graphs, we derive the parameters such that the mean total number of connections $N_c$ in the graph is conserved. Table II summarizes the numerical values chosen for the corresponding model parameters.

*a. Erdős-Rényi (ER).* In this network model [44], every connection has the same probability of being realized: $p = \frac{N_c}{N_e}$, where $N_e$ is the total number of possible connections in the graph. Note that this network model maximizes the entropy under the constraint that the mean number of connections is fixed [45].

*b. Directed configuration model (DCM).* In a directed configuration model [46], a two-step probabilistic process is applied. First, random indegrees and outdegrees are drawn for each node such that the total number of connections across nodes is preserved (we fix these numbers for all graph instances). Second, connections are established by randomly matching each outgoing connection with an incoming connection. Thus, two nodes can have more than one connection, and the resulting adjacency matrix is not strictly binary.

*c. One-cluster Erdős-Rényi (OC).* Based on an Erdős-Rényi (ER) graph, we introduce a single cluster by increasing $p$ between a certain subset of nodes of the network, while uniformly decreasing $p$ for all other connections such that $N_c$ is conserved. The relative increase of $p$ is denoted by $r$, and the location of the cluster on the diagonal is defined by the bounding indices $b^{\text{lower}}$ and $b^{\text{upper}}$.

*d. Two-cluster Erdős-Rényi (TC).* For the two-cluster ER network, we create two nonoverlapping clusters on the diagonal using the same method as in the OC model. The nodes that form the clusters are chosen such that there is maximal overlap with the single cluster of the OC model: the first cluster starts at the same index $b^{\text{lower}}$ and extends up to index $b^{\text{mid}}$, and the second cluster starts at index $b^{\text{mid}} + 1$ and extends up to index $b^{\text{upper}}$.

*e. Watts-Strogatz (WS).* We create a small-world network following [47]. Here $N_n$ nodes initially form a ring, where each node is connected to $k = \frac{N_c}{N_e}(N_n - 1)$ of its nearest neighbors. Afterwards, all connections are uniformly redistributed with probability $p_{WS}$. Note that this model is undirected.

*f. Barabási-Albert (BA).* As an example of a scale-free network, we create Barabási-Albert networks as introduced in [48]. Here, from an initial star graph with $m = \frac{N_c}{N_e}(N_n - 1)/2$ nodes, new nodes are added subsequently until the desired number of nodes $N_n$ is reached. Each added node is connected to $m$ existing nodes, where the probability of each existing node being selected for a new connection is proportional to the number of connections it already has. Note that this model is undirected.

### D. Brain data

We apply SAS to compare nonsquare matrices of brain activity in response to visual stimuli. We use an openly available data set, which has an extensive description of the task and recording apparatus [49]. In the experiments, the activity of neurons in the primary visual cortex (V1) of one macaque monkey (*Macaca mulatta*) was recorded using several extracellular electrode arrays (Utah arrays, $8 \times 8$ electrodes). The quality of the signals was assessed based on the signal-to-noise ratio and channel impedance. For details of the data recording and processing we refer to [49]. Here we focus on a single array (ID = 11) during a receptive field mapping

task. In this task, for each trial the macaque had to fixate its gaze to the center of the screen for 200 ms. Subsequently, one bright bar moved across the screen for 1000 ms in one of four directions: rightward (R), leftward (L), upward (U), or downward (D). The different task modalities are in the following referred to as trial types. For each trial type, there are $N = 120$ repetitions.

The activity time series recorded from the electrodes was processed to obtain the multi-unit activity envelope (MUAe) with a sampling rate of 1 kHz, a commonly used signal as a proxy for neuronal firing rates [50]; see [51] for specific details of the processing. We align the trials to the peak response, defined as the maximum response from the average MUAe across electrodes, and cut data in a window ±200 ms around the alignment trigger. This yields one $64 \times 400$ matrix per trial: 64 electrodes during 400 ms at 1 kHz; see examples in Fig. 7(b). We then group the matrices by trial type for comparison by SAS.

### E. Degenerate matrices

To assess SAS in the presence of degenerate singular values we construct matrices with various levels of degeneracy from random matrices and random graphs. Given a matrix $M = U\Sigma V^T$, we introduce a degeneracy parameter $d$ forcing

$$\sigma^i = \cdots = \sigma^{i+d} = \frac{1}{d}\sum_{j=i}^{i+d}\sigma^j \tag{14}$$

and define the degenerate matrix as

$$M^d = U\Sigma^d V^T, \tag{15}$$

where in $\Sigma^d$ the corresponding singular values are replaced by their empirical mean. We distinguish three special cases of starting index $i$ such that either the highest (first), lowest (last), or central singular value are first made degenerate, referred to as *H*, *L*, and *C degenerate*, respectively. In the C-degenerate case, $i$ is adjusted such that the central singular value lies in the center of the interval $[i, i+d]$. Finally, we denote the relative degeneracy of an $N \times N$ matrix as $D = d/N$.

## III. RESULTS

We present a measure for assessing the structural similarity between two arbitrary, real $m \times n$ matrices $M_a, M_b$ named *singular angle similarity* (SAS). The measure is based on singular value decomposition (SVD), which introduces the left and right singular vectors with corresponding singular values [Eq. (1)]. SAS exhibits the following properties:

(1) SAS attains values between 0 and 1 where higher values imply greater similarity.

(2) SAS is invariant under actions of identical orthogonal maps from the left or the right on the compared matrices; this includes the consistent permutation of rows and columns as a special case (Sec. S1.2 [37]).

(3) SAS is invariant under transposition of both matrices (Sec. S1.2 [37]).

(4) SAS is invariant under scaling with a positive factor; in particular, SAS = 1 for $M_b = c_1 M_a$, $c_1 \in \mathbb{R}^+$ (Sec. S1.3 [37]).

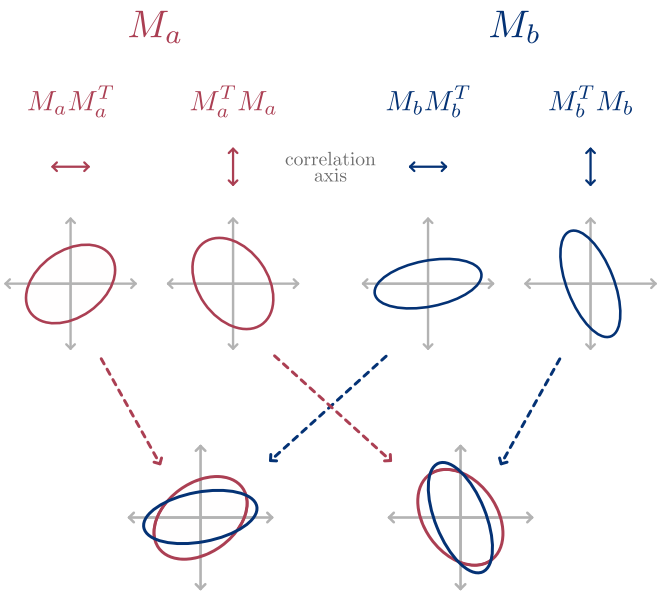


FIG. 2. Geometric interpretation of SAS. Matrices $M_a$ (red) and $M_b$ (blue) as in Fig. 1. The eigenvectors of $M_iM_i^T$ and $M_i^TM_i$ (same colors) scaled by the square root of their eigenvalues span the main axes of ellipsoids. These square matrices capture the correlation structure of $M_i$ along the horizontal and vertical axis, respectively (double-headed colored arrows). SAS compares the angles between the corresponding ellipsoids (dashed colored arrows).

(5) SAS is zero if the two compared matrices are equal up to a negative factor: SAS = 0 for $M_b = c_2 M_a$, $c_2 \in \mathbb{R}^-$ (Sec. S1.3 [37]).

(6) If $w(x, y) = \sqrt{xy}$, SAS is invariant under isotropic scaling [6]

Thus, SAS predominantly highlights structural differences between the matrices. The derivation of this measure is presented in Sec. II A.

### A. Geometric interpretation of SAS

Singular angle similarity has a geometric interpretation. The left and right singular vectors of $M_i$ ($i \in \{a, b\}$) are the respective eigenvectors of the square matrices $M_iM_i^T$ and $M_i^TM_i$. Further, $M_iM_i^T$ and $M_i^TM_i$ have the same eigenvalues (the squared singular values of $M_i$). Consider for each of these symmetric positive-definite matrices a hyperellipsoid spanned by the respective eigenvectors scaled by their eigenvalues (Fig. 2).

The hyperellipsoid collapses in most dimensions as matrices $M_i$ typically have only a small number of large singular values (cf. [52]). Dimensions associated with the largest singular values dominate its shape, and the angle between the corresponding left and right singular vectors of the matrices $M_a$ and $M_b$ are of main relevance for SAS. Thus, a high SAS indicates that the hyperellipsoids are aligned, whereas a low SAS indicates misalignment or different shapes. If two matrices share two-dimensional structural features, their hyperellipsoids will be similarly shaped and point into similar directions, producing a high SAS. $M_iM_i^T$ and $M_i^TM_i$ are the correlation matrices up to a normalization by the number of rows and columns, respectively, and the subtraction of the

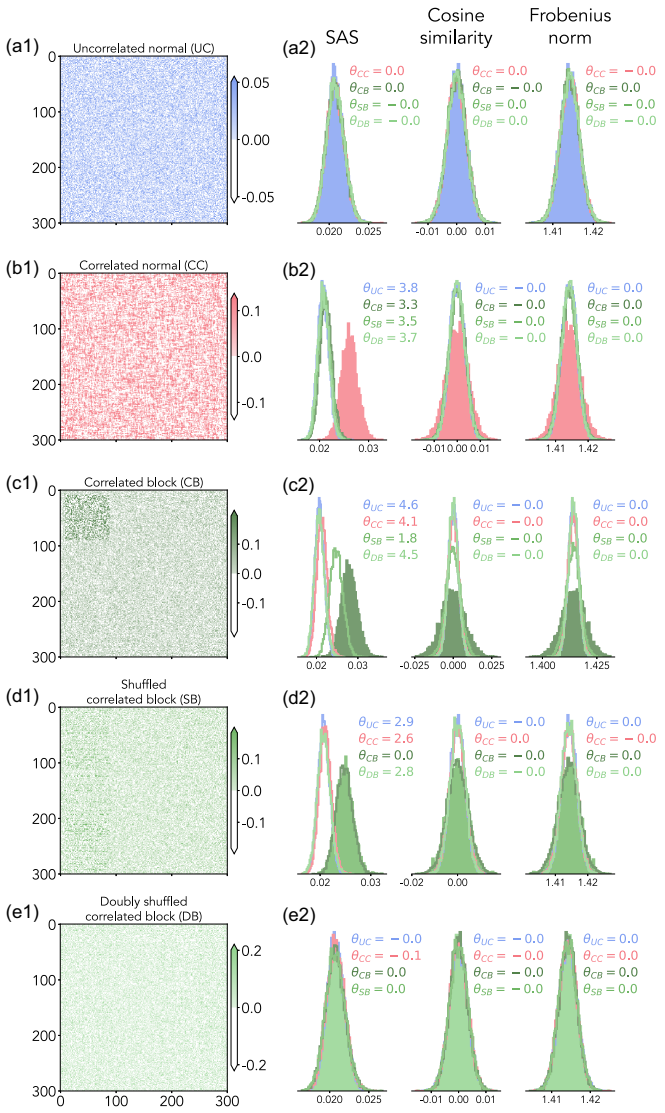


FIG. 3. Comparison of SAS with standard measures. [(a1)–(e1)] Single instances of the different random matrix classes. For visibility, negative values are shown as white. [(a2)–(e2)] Histograms of SAS, cosine similarity, and the Frobenius norm between instances of the random matrix classes ($n = 100$, all pairs compared). Filled distributions indicate self-similarities (self-distances), nonfilled ones indicate cross-similarities (cross-distances). Legends show effect sizes $\theta$ for the comparison between distributions.

mean. Therefore, SAS takes into account the correlation structure along both axes of the matrices. This distinguishes the measure from common methods such as cosine similarity and the Frobenius norm.

### B. Comparison with standard measures for random matrices

By its very definition, SAS captures two-dimensional structures that are invisible to traditional measures of matrix similarity. Figure 3 shows the ability of different measures to discriminate between classes of random matrices with such structure.

Concretely, we compare SAS with cosine similarity,

$$\langle M_a, M_b\rangle_F = \mathrm{tr}\big(M_a M_b^T\big), \tag{16}$$

and Frobenius distance,

$$||M_a - M_b||_F = \sqrt{\langle M_a - M_b, M_a - M_b\rangle_F}, \tag{17}$$

where we normalize the matrices such that

$$||M_a||_F = ||M_b||_F = 1.$$

Figures 3(a1)–3(e1) show single instances of the five matrix classes defined in Sec. II B. We first calculate the *self-similarity*, the pairwise similarity between instances of the same random matrix class (excluding self-comparisons), and the *cross-similarity*, which refers to the similarity between instances of different classes [Figs. 3(a2)–3(e2)]. We analogously define self- and cross-distance for the Frobenius distance. Subsequently, we investigate whether the different measures distinguish the random matrix classes from each other based on realizations of their particular structures. Fundamentally, this works only if the structure quantified by a measure is more similar between instances of the same class than across classes. Thus, for SAS and cosine similarity, the self-similarity must be meaningfully greater than the cross-similarities. Conversely, since the Frobenius norm measures a distance rather than a similarity, the self-distance must be smaller than the cross-distance. We call a difference meaningful if the effect size $\theta$ between pairs of distributions is greater than one. Assuming an underlying Gaussian model for the distributions, we employ the definition

$$\theta = \frac{\mu_{\mathrm{self}} - \mu_{\mathrm{cross}}}{\sqrt{\frac{\sigma^2_{\mathrm{self}} + \sigma^2_{\mathrm{cross}}}{2}}} \tag{18}$$

of the "Cohen's D" effect size [53] underlying the common Student's and Welch's $t$ statistics [54,55], where $\mu$ and $\sigma$ are the mean and standard deviation of the self- and cross-similarity distributions. Thus, two distributions are meaningfully different if the distance of their means is greater than the quadratic mean of their standard deviations.

Figure 3(a2) shows that UC matrices cannot be distinguished from the other matrix classes by any measure. This is expected: since the entries are independent, there is no detectable structure. In particular, this means that no structure is shared between different UC matrices or between UC matrices and matrices of other classes. Geometrically, this corresponds to ellipsoids that are oriented in random directions for each instance.

By definition, CC matrices exhibit shared fluctuations that induce similarity between different instances of the matrices. However, cosine similarity and the Frobenius norm fail to identify the common correlation structure [Fig. 3(b2)]. Only SAS separates the self- and cross-similarity meaningfully and can thereby distinguish this matrix class from the others.

A similar conclusion holds true for CB matrices, where the correlated structure is embedded into an otherwise uncorrelated matrix [Fig. 3(c2)]: again, only SAS separates the self- and cross-similarities meaningfully.

Next, we consider SB matrices. By construction, these are CB matrices with permuted rows. Between SB matrices, the CB correlation structure along the horizontal axis (quantified

by $MM^T$) is destroyed while the correlation structure along the vertical axis (quantified by $M^TM$) stays the same. Since SAS takes into account both, it detects similarity between SB and CB matrices despite the permutation of the rows. This leads to a higher cross-similarity between CB and SB matrices than between CB and the other matrix classes [Fig. 3(c2)]. Since the block structure exhibited by CB matrices can be viewed as one specific permutation of the rows, the self-similarity of SB and the cross-similarity between SB and CB follow the same distribution [Fig. 3(d2)] while SB matrices are separable from UC and CC matrices. The cosine similarity and the Frobenius norm fail to separate SB matrices from the other classes. The choice of the axis along which we permute is arbitrary; the results are identical if we permute columns instead of rows.

Finally, we turn to DB matrices [Fig. 3(e2)]. Here, starting from CB matrices, both the columns and rows are permuted. Consequently, between DB matrices a correlation structure is neither retained along the horizontal nor along the vertical axis. Neither SAS nor the other measures can detect similarity between DB matrices in comparison with DB matrices and matrices of the other classes.

Why are these examples relevant? They show that SAS captures certain two-dimensional correlation structures between instances. In contrast, the traditional measures cannot identify them. Additionally, SAS retains similarity even after permutation along one axis—including shifts as a special case. This is relevant in practical applications, for example in the analysis of highly parallel time series: even if the time series are not aligned, SAS exposes structural similarities. However, if both rows and columns are randomly permuted SAS fails to identify similarity. This is expected since in this case there is no two-dimensional correlation structure between instances left for SAS to identify.

In addition to the metrics described above, we also test SAS against linear CKA and the angular Procrustes similarity (Sec. S1.4 [37]). While for these measures CC and CB matrices can be better separated, this does not generalize to SB matrices. Moreover, the separations suffer from spurious high similarity for UC matrices, violating the expected hierarchy of similarity in multiple cases (Sec. S1.4 [37]).

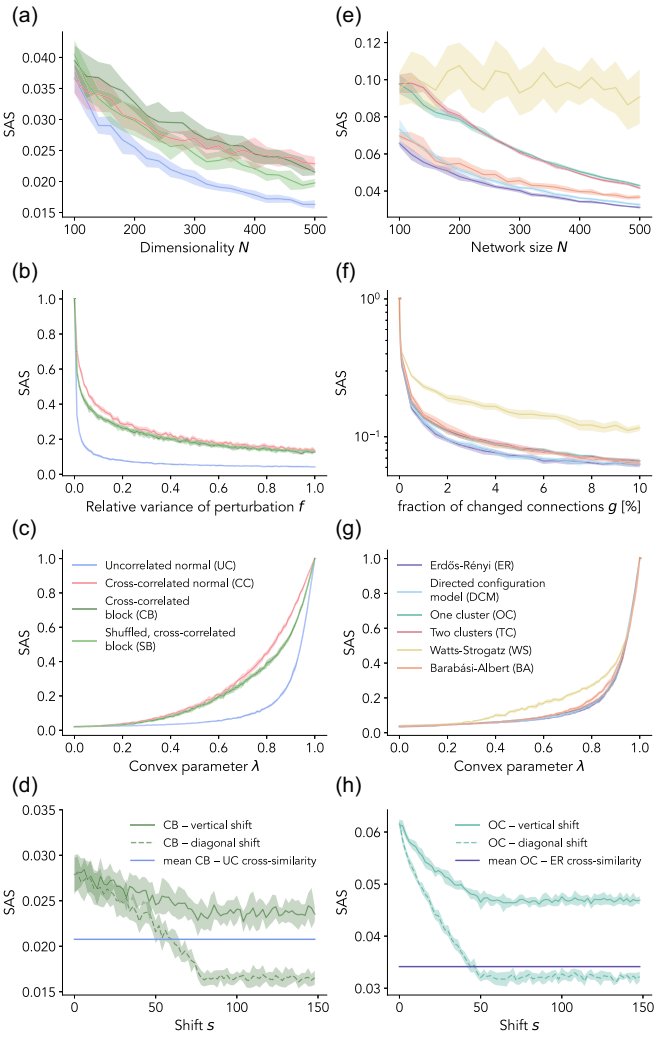


FIG. 4. Characterization of SAS. In all panels, lines indicate the mean and shadings indicate the standard deviation over 10 realizations. In cases of constant matrix size, $N = 300$. Panels (a)–(c) and (e)–(g) share the same legends. (a) Self-similarity for varying dimensionality $N$. (b) SAS between identical matrix instances for varying variance of an additive perturbation $\sigma^2_{\text{pert}} = f\sigma^2$. (c) SAS between matrices when increasing the degree of putative structural similarity quantified by $\lambda$. (d) SAS between CB matrices where the correlated block is shifted either vertically or diagonally for one matrix. (e) Self-similarity for varying network size $N$. (f) SAS between identical network model instances where the number of individual connections that are changed is gradually increased. (g) SAS between network model instances when increasing the degree of putative structural similarity quantified by $\lambda$. (h) SAS between OC matrices where the cluster is shifted either vertically or diagonally for one matrix.

## C. Characterization

### 1. *Scale dependence and robustness*

To assess the dependence of SAS on matrix size, we calculate the self-similarity for increasing dimensionality $N$ and observe a decreasing SAS for all random matrix classes [Fig. 4(a)]. Since the probability distribution for an angle between two random vectors increasingly centers around $\pi/2$ with increasing dimensionality [56], the resulting SAS between UC matrices decreases with increasing matrix dimensionality. This intuition generalizes to the other matrix classes. Hence, a quantitative comparison of SAS values is only reasonable for matrices of the same dimensionality.

Next, we investigate how SAS decreases between two identical copies of a matrix when gradually perturbing one of them. We analytically study SAS between a matrix $M$ and a perturbed version of itself, $M + \sqrt{\epsilon}W$, using Rayleigh-Schrödinger perturbation theory [57]. Here $\sqrt{\epsilon}$ is chosen as a perturbation parameter since this ensures a linear scaling of the variance of the perturbation matrix with $\epsilon$. We find that, for a large class of perturbations, SAS follows $1 - \arccos[1 - O(\epsilon)]/\frac{\pi}{2}$ (see Sec. S1.5 [37]). This implies that small differences are identified as dissimilarities arbitrarily fast ($\frac{d\arccos(1-x)}{dx} \to \infty$ for $x \to 0$). Thus, SAS is sensitive to small differences in the compared matrices. For an empirical analysis, we study the sensitivity of SAS under perturbations of additive noise of the form $\tilde{M} = M + W$ where

$W$ is a matrix of the UC class with zero mean and variance $\sigma^2_{\text{pert}} = f\sigma^2$, $0 \leqslant f \leqslant 1$. As predicted analytically, Fig. 4(b) shows a rapid fall-off for small perturbations that continues as a gradual decrease for each of the considered classes.

Next, we numerically study SAS while adding structure to a noise matrix. In particular, we calculate SAS between a matrix $M$ and the convex combination of the same matrix with a UC matrix $N$:

$$(1-\lambda)N + \lambda M \quad \text{for } \lambda \in [0,1]. \tag{19}$$

Figure 4(c) shows that SAS is well behaved also for small values of $\lambda$: it increases smoothly when adding structure for all matrix classes.

Finally, starting from CB matrices, we investigate the change of cross-similarity when shifting the correlated block of one matrix either vertically or diagonally by $s$ indices [Fig. 4(d)]. In both cases, SAS gradually decreases and saturates once the blocks are nonoverlapping ($s = 80$). Importantly, SAS saturates to a value higher than the mean CB-UC cross-similarity for the vertical shift, whereas it saturates to a value lower than that for the diagonal shift. Thus, SAS identifies shared structure when it is shifted vertically, but not when it is shifted diagonally. While potentially counterintuitive, this can be understood when considering the case of DB matrices [Fig. 3(e2)]: the diagonal shift is a specific example of a permutation in both directions, and once the blocks are nonoverlapping, there is no correlation structure between the two matrices that SAS can identify. We conclude that SAS cannot detect similarity between matrices if a sufficient amount of the relevant structure is moved across instances.

We perform an analogous analysis for network adjacency matrices of six different graph models: Erdős-Rényi (ER), directed configuration model (DCM), one cluster (OC), two clusters (TC), Watts-Strogatz (WS), and Barabási-Albert (BA), as defined in Sec. II C. The results are qualitatively similar to those obtained for the four classes of random matrices: SAS decreases with increasing network size $N$ [Fig. 4(e)], SAS rapidly decreases for small perturbations [Fig. 4(f)], SAS gradually increases when adding structure [Fig. 4(g)], and SAS decreases when shifting either vertically or diagonally [Fig. 4(h)]. A notable exception is that for WS networks, SAS does not decrease when increasing $N$. In this network model the number of nearest neighbors each node is connected to scales with $N$. Therefore the correlation in the adjacency matrix also scales with $N$, rendering the similarity measured by SAS independent of $N$. For investigating the effect of a perturbation on identical network matrices [Fig. 4(g)], we define the gradual change such that an increasing fraction $g$ of matrix elements is altered. Specifically, for each of the $gN^2$ randomly selected matrix elements, existing connections are removed and missing connections are established with a multiplicity of one. For the binary adjacency matrices this corresponds to bit flipping the corresponding entries. In the case of adding structure [Fig. 4(g)], we choose the ER adjacency matrices as the noise component $N$. Note that while the sum over all entries stay the same on average under the convex combination, the entries are not confined to natural numbers anymore.

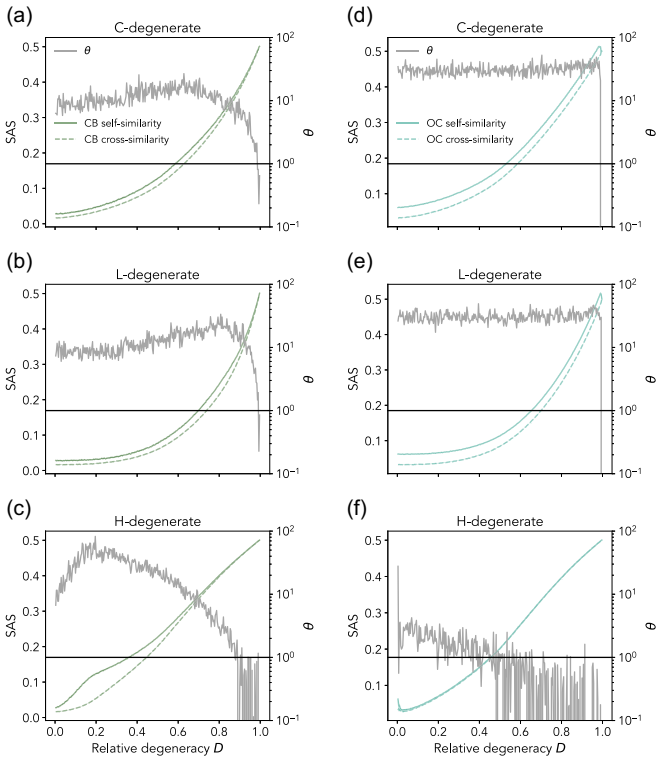


FIG. 5. SAS for degenerate matrices. In all panels, lines indicate the mean, the shading the standard deviation over 10 realizations. Solid colored lines show self-similarity between CB (OC) matrices with identical parameters (the same as in Tables I and II). Dashed colored lines show cross-similarity between two CB (OC) matrices where the second matrix exhibits a block (cluster) shifted to $b^{\text{lower}} = 210$ and $b^{\text{upper}} = 290$ ($b^{\text{lower}} = 200$ and $b^{\text{upper}} = 250$). Gray lines indicate effect size $\theta$ (right axes). In all panels, the fraction $D$ of degenerate singular values is varied. $N = 300$. (a), (d) C-degenerate matrices, where the singular values are made degenerate starting from the center singular value. (b), (e) L-degenerate matrices, where the singular values are made degenerate starting from the lowest singular value. (c), (f) H-degenerate matrices, where the singular values are made degenerate starting from the highest singular value.

### 2. *Degenerate singular values*

We evaluate the discriminability of matrix classes with SAS in the presence of degenerate singular values (Fig. 5). Specifically, we quantify how well CB matrices with nonoverlapping blocks and OC matrices with nonoverlapping clusters can be discriminated when making the singular value spectrum degenerate. We introduce degeneracy gradually as described in Sec. II E.

In the case of CB matrices, for which the singular value spectrum exhibits a smooth decay (not shown), SAS is able to distinguish matrices with blocks at different positions even for high degeneracy ($\theta > 1$ up to $D \approx 0.85$) as shown in Figs. 5(a)–5(c). Here the effect size initially increases with increasing H-degeneracy even though the destroyed structure corresponds to the singular angles with the highest weighting in the calculation of SAS.

In contrast, the spectrum of singular values of OC matrices is dominated by one singular value (not shown), which is of highest relevance for SAS. When introducing

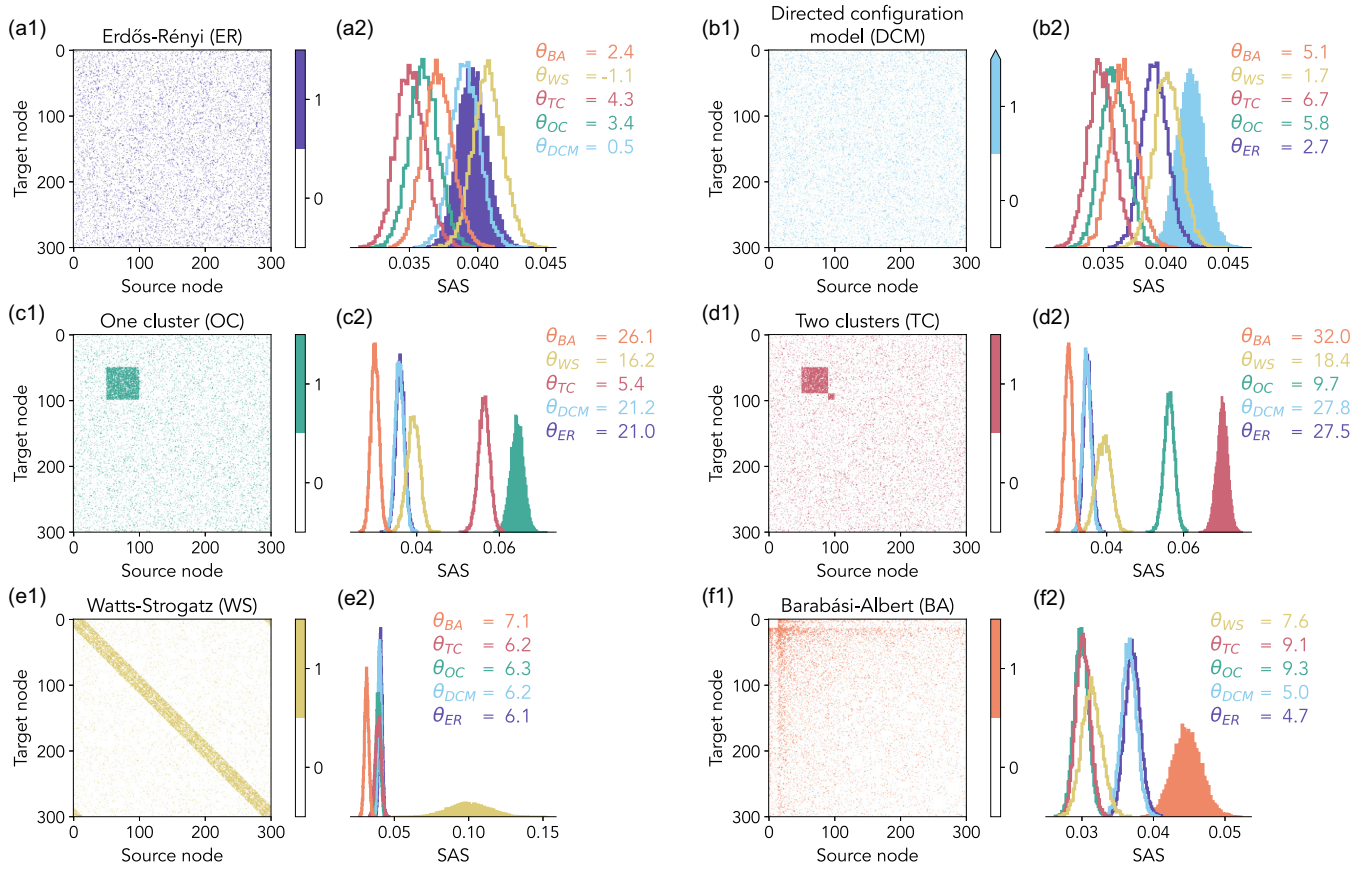


FIG. 6. Self- and cross-similarity of different network models. [(a1)–(f1)] Single instances of the different network models. Colored matrix elements indicate a connection between nodes. For the DCM model, connections with a multiplicity higher than one are shown with the same color intensity as single connections. [(a2)–(f2)] Histograms of SAS between instances of the network models ($n = 100$, all pairs compared). Filled distributions indicates self-similarities, nonfilled ones indicate cross-similarities. Legends show effect sizes $\theta$ for the comparison between self- and cross-similarities.

degeneracy from the center [Fig. 5(d)] or starting from the smallest singular value [Fig. 5(e)], $\theta$ remains roughly constant until $d = 299$, $d = 300$, respectively. At precisely these values, the largest singular value becomes degenerate and $\theta$ drops below 1. For the left-degenerate case [Fig. 5(f)], some discriminability is retained for small values of $d$ even though the largest singular value is made degenerate already at $d = 2$.

This implies that, in a certain range depending on the use case, SAS is robust against the perturbation of large outliers in the distribution of singular values. Moreover, these cases suggest that SAS is not principally limited by the presence of repeated singular values.

### D. Categorization of random graphs

We apply SAS to the adjacency matrices that describe the network architecture in various directed and undirected probabilistic graphs as defined in Sec. II C. Example adjacency matrices of network instances are shown in Figs. 6(a1)–6(f1). Since these matrices $M_i$ only contain non-negative entries, so do $M_i M_i^T$ and $M_i^T M_i$. The Perron-Frobenius theorem [58] guarantees that the left and right singular vectors corresponding to the largest singular value have only non-negative or only nonpositive entries (cf. Sec. S1.6 [37]). As such, these singular vectors are confined to a single orthant of the $N$-dimensional vector space. Even if these vectors are random, they cannot be assumed to be orthogonal. Indeed, for the ER network, for which all other singular vectors are of random orientation, the first left and right singular vectors scatter around the vector $(1/\sqrt{N}, \ldots, 1/\sqrt{N})^T$ across instances. Therefore, the first singular vectors necessarily enclose smaller angles across models compared to the other pairs of singular vectors. Consequently, information regarding the difference between models—which is encoded most strongly in the first singular vectors—is reduced. Thus, it is *a priori* not clear whether SAS reliably distinguishes between network models.

*a. Self-similarity of network models.* First, we examine the self-similarity of the network models [Figs. 6(a2)–6(f2)]. We find that ER networks exhibit the lowest self-similarity compared to all other network models. This is consistent with the ER network model maximizing the entropy under the constraint that the average number of connections is constant: ER networks have the least structure that is shared across instances. This can be also understood from their definition inasmuch as each connection is realized independently with the same probability. In this sense, ER networks are analogous to the UC random matrices. The other network models instead feature structural properties that are consistent across instances, stemming from shared variations in the connection probability. This is most obvious for the OC and TC network

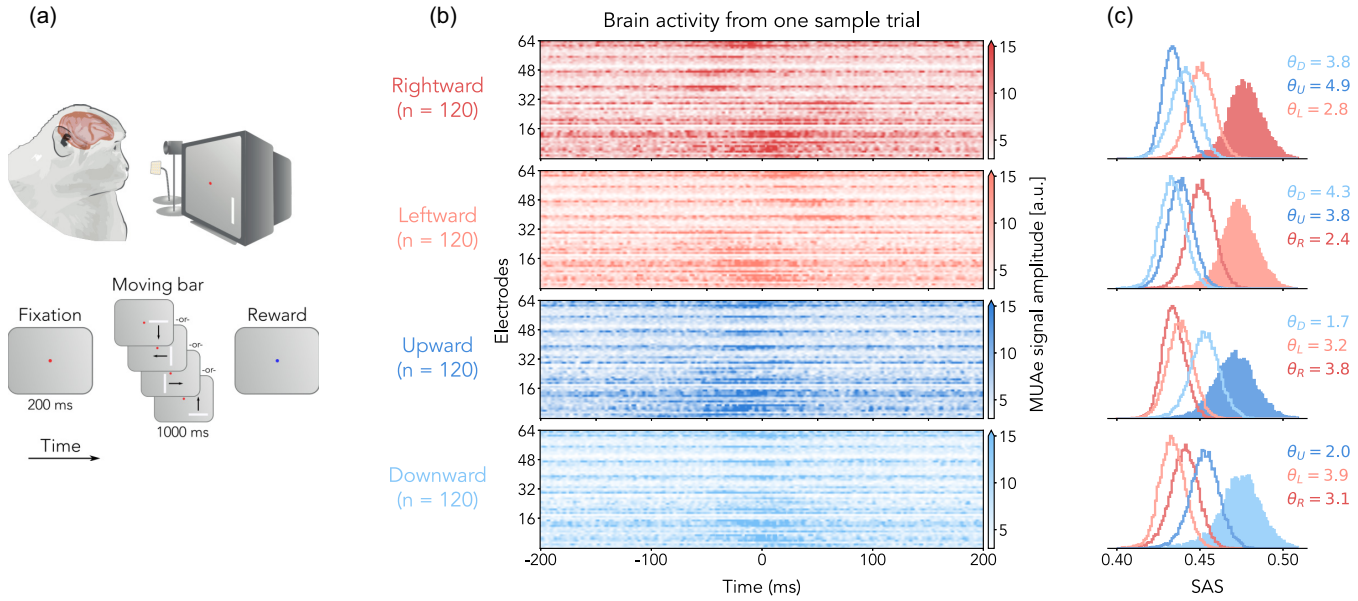


FIG. 7. Comparing nonsquare matrices of brain activity with SAS. (a) Schematic diagram of the neuroscience experiment. (b) Single-trial brain activity at identical recording electrodes (64 electrodes, vertical) for each type as example (trial types distinguished by color). Larger values of matrix entries indicate higher MUAe activity (shading in color bar, arbitrary units). (c) Histograms of SAS between all ($n = 120$) individual trials. Filled distributions [color code as in (b)] indicate self-similarities, nonfilled ones indicate cross-similarities. Legends show effect sizes $\theta$ for the comparison between self- and cross-similarities.

models (analogous to the CB random matrix class), where certain subgroups of nodes have a higher connection probability $p$ among themselves compared to the rest of the network. Further, we expect WS networks to have reliably detectable structure, i.e., high self-similarity, as every node has dominant local connectivity. SAS confirms these expectations, as seen when comparing the respective self-similarity distributions in Figs. 6(a2)–6(f2).

*b. Self-similarity vs cross-similarity.* Second, we study whether SAS can differentiate between the particular structures present in the adjacency matrices of the network models. Using the effect size defined in Eq. (18), Figs. 6(a2)–6(f2) confirm that the self-similarity is meaningfully higher than the cross-similarity except for special cases.

First, ER networks do not exhibit $\theta > 1$ for all cases. As for the UC matrices, this is expected as all matrix elements are uncorrelated. As a matter of fact, the cross-similarity with WS networks yields even higher SAS values than the self-similarity of ER. Why is this the case? The first left and right singular vectors of both ER and WS networks scatter around $(1/\sqrt{N}, \ldots, 1/\sqrt{N})^T$. The deviation between singular vectors of ER networks, however, is larger than that between those of WS networks across instances. This leads to a better alignment, i.e., a higher angular similarity, of the singular vectors, resulting in a higher SAS between ER and WS as compared to ER and ER networks.

Second, we note that SAS distinguishes between the OC and TC networks despite overlapping clusters. Figures 6(c2) and 6(d2) show that the respective self-similarities are closer to the cross-similarity of OC and TC than to the other cross-similarities. Thus, SAS identifies the clustered networks to be more similar among each other than compared to the remaining networks.

We conclude that SAS is sensitive to the structure present in matrices, enabling it to distinguish between model classes. The same conclusion also holds true for nonsquare matrices of network connectivity where a full graph is instantiated, but only subsamples are analyzed with SAS (see Sec. S1.7 [37]).

### E. Separation of brain states

We investigate brain activity originating from different experimental trials as a use case for SAS with nonsquare matrices of experimental data. The publicly available data set from [49] is based on extracellular recordings from the visual cortex of a macaque monkey. In the experiment, bright bars move across a screen in one of four directions [rightward (R), leftward (L), upward (U) or downward (D)], evoking a strong neural response [Fig. 7(a)]. The data consists of the multi-unit activity envelope (see Sec. II) yielding one $64 \times 400$ matrix per trial; sample matrices are shown in Fig. 7(b).

Neurons in the primary visual cortex (V1) respond according to their feature selectivity, primarily stimulus location [59] and orientation [60], but also movement direction [61]. Beyond these well-known response properties of single neurons, the population activity—represented as a two-dimensional spatiotemporal matrix—may reveal additional information about brain dynamics. By applying SAS, we investigate shared variability across both time and neurons.

In the data set at hand, SAS reveals that neural activity of all trial types exhibits higher self- than cross-similarity with effect sizes $\theta > 1$ [Fig. 7(c)]. Trials with stimulus movement along the same axis (L-R or U-D) are more similar to each other than to ones with orthogonal stimulus movement. This is a desirable outcome of SAS: in L-R (resp. U-D) trials, neurons that share an orientation tuning aligned with the stimulus are

expected to have a higher probability of a strong response. The consideration implies that the shared orientation of the bar stimulus in L and R (resp. U and D) trials leads to more similar responses in these trial pairs.

In this use case, SAS outperforms common measures of matrix similarity. Both cosine similarity and the Frobenius norm can distinguish trial types—albeit with lower $|\theta|$ than SAS—but fail to identify the shared variability along the same axis (L-R or U-D). Symmetric CKA and the angular Procrustes similarity fail to reliable distinguish trials in most cases (Fig. S3 [37]). Therefore, the use case highlights the ability of SAS to identify the two-dimensional structure of matrices in experimental data.

As an essentially linear measure operating directly on the data, SAS might suffer in the presence of strong nonstationarities and highly nonlinear trajectories in the phase spaces. Ostrow *et al.* [62] suggest to overcome this problem by representing the nonlinear system as a high-dimensional linear one using dynamic mode decomposition [63]. In a second step, they then assess the similarities of the linear representations via a Procrustes-style minimization problem using base changes instead of actions of orthogonal matrices. Guilhot *et al.* [64] show that the method can identify the emergence of representations during learning in recurrent networks —a task at which static measures like CKA or a Procrustes-based approach fail. It will be interesting to see how the assessment of similarity changes if SAS is used instead of the similarity measure employed by the authors in [62] for comparing the linear high-dimensional representations of the dynamical systems.

## IV. DISCUSSION

Here we present *singular angle similarity* (SAS): a method for comparing the similarity of real matrices of identical shape based on singular value decomposition. SAS is invariant under transposition and consistent relabeling of coordinates, and reliably detects structural similarities in the compared matrices. It generalizes the angular similarity (Sec. S1.8 [37]) for vectors and the eigenangle score for symmetric positive definite matrices by Gutzen *et al.* [33]. For a particular choice of weight function and vector similarity measure, SAS and the eigenangle score are equivalent for positive-definite symmetric matrices. Gutzen *et al.* [33] introduce an extension of the eigenangle score for asymmetric adjacency matrices of certain network types by choosing a specific analytical mapping of the complex-valued eigenvalues and vectors: shift the real part of the eigenvalues by the spectral radius of the matrix, and calculate the Euclidean angle between the complex eigenvectors [65]. However, this choice is not unique—other mappings are also possible. In addition, the resulting similarity measure is not invariant under joint transposition of the compared matrices. In SAS, the singular vectors are real-valued, the singular values are non-negative, and the measure is transposition invariant. Thereby, it circumvents these limitations and admits a more natural generalization to nonsymmetric and nonsquare matrices. We here choose angular similarity as the vector similarity measure, and a specific weight function for the resulting angles [Eq. (8)]. Other choices are possible, such as cosine similarity for the former, making SAS easily adaptable. By definition, SAS can only be applied to matrices with identical shape. A possible extension that allows one to apply SAS to matrices of different shape is sketched in Sec. II A.

*a. Interpretability.* Since SAS is sensitive to matrix size (Sec. III C) it can only be interpreted in relative terms. A quantitative reference can be obtained by choosing a use-case-specific null model (e.g., Erdős-Rényi for network models) for which the corresponding distribution of SAS can be determined numerically. Similarity as indicated by SAS can then be interpreted in terms of this baseline. Since realizations of single matrices exhibit fluctuations, evaluating SAS from a single observation may be misleading. Instead, one should consider the SAS distribution of an ensemble of realizations when possible. Such a distribution can then be interpreted with respect to the reference distribution obtained with the null model by means of an effect size [Eq. (18)] or a statistical two-sample test of choice. Given the broad range of potential use cases, analytical descriptions of a null distribution and associated calculations of $p$ values as in [33] seem difficult or even impossible for many cases, and we suggest an empirical approach based on explicit choices of null models as described above.

However, we highlight that an empirical approach to a statistical assessment of methods based on the analyses of matrices is not always required. Indeed, some methods for which key quantities are related to singular values or eigenvalues allow for analytical descriptions of statistical tests or other assessments (e.g., [66–70]). Even though the overlap between eigenvectors has also been investigated analytically in special cases [71], focusing on singular angles makes a fully mathematical derivation of significance tests in many cases impossible to the best of our knowledge. We hence advocate for the empirical approach outlined above.

*b. Limitations.* While SAS generally highlights structural features in matrices stemming from their correlations along rows and columns, it also suffers from shortcomings in certain situations. In the presence of sufficiently strong noise, the order of singular values may change even if two matrices encode the same underlying information. This leads to a different pairing of singular vectors when computing SAS, resulting in a low score even though the matrices result from a common construction process. This can be partially be addressed by rounding singular values, thereby potentially introducing degeneracy. This is treated in SAS by deriving singular angles from the canonical angles between the degenerate subspaces. If the degenerate subspaces are large, SAS may lose its sensitivity: in the extreme case of fully degenerate matrices where the degenerate subspace is the full space, no discrimination is possible using SAS. Additionally, if the matrices under consideration have few strongly dominating singular values, and these are degenerate, the discriminability using SAS may suffer. Further, SAS cannot detect similarity between matrices if rows and columns are inconsistently permuted. This includes shifts along both axes as a special case. Hence, SAS cannot be used when translation invariance is desired, for instance, in the detection of objects in images independent of their location.

*c. Applications.* Beyond network connectivity or brain activity matrices, potential applications include the analysis of nonsymmetric matrices obtained with measures for the flow

of information. Examples of such measures include, but are not limited to, Granger causality [72], or transfer entropy [73] (also see [74] for a machine learning inspired view on the topic). Additionally, SAS can help assess similarity in more classical settings, e.g., when studying cross-covariances.

In conclusion, SAS can be used to analyze the structural similarity of any real-valued data that can be represented in matrix form. Such data can come from any field of research. Coupled with domain knowledge, SAS may reveal hidden structures in the data, supporting existing methodologies and enabling new insights.

## ACKNOWLEDGMENTS

This work has been supported by NeuroSys as part of the initiative "Clusters4Future" by the Federal Ministry of Education and Research BMBF (03ZU1106CB); the DFG Priority Program (SPP 2041 "Computational Connectomics"); the EU's Horizon 2020 Framework Grant Agreement No. 945539 (Human Brain Project SGA3); the European Union's Horizon Europe Programme under the Specific Grant Agreement No. 101147319 (EBRAINS 2.0 Project); the Ministry of Culture and Science of the State of North Rhine-Westphalia, Germany (NRW-network "iBehave," Grant No. NW21-049); and the Joint Lab "Supercomputing and Modeling for the Human Brain." The authors thank the two anonymous reviewers for their constructive remarks and helpful suggestions that greatly improved the quality of the manuscript.

Author contributions: conceptualization: J.A., A.C.K., R.G.; methodology: A.C.K., J.A., R.G.; software: J.A., R.G., A.C.K., A.M.-G.; formal analysis: J.A., A.C.K., R.G., A.M.-G.; writing–original draft: J.A., A.C.K., R.G., A.M.-G.; writing–review and editing: all; visualization: J.A., A.C.K., A.M.-G., R.G.; supervision: M.De., S.v.A., M.Di.; funding acquisition: S.G., S.v.A., M.Di.

## DATA AVAILABILITY

Code to calculate singular angle similarity (SAS) is openly available on GitHub [75] and Zenodo [76], and in the validation test library NetworkUnit [77,78]. The data and code to reproduce the results from this paper can be found on Zenodo [79].

[1] M. E. Newman, The structure and function of complex networks, SIAM Rev. **45**, 167 (2003).

[2] B. Calsbeek and C. J. Goodnight, Empirical comparison of G matrix test statistics: Finding biologically relevant change, Evolution **63**, 2627 (2009).

[3] R. V. Solé and M. Montoya, Complexity and fragility in ecological networks, Proc. R. Soc. Lond. B **268**, 2039 (2001).

[4] C. Piccardi, L. Calatroni, and F. Bertoni, Clustering financial time series by network community analysis, Int. J. Mod. Phys. C **22**, 35 (2011).

[5] A. Morcos, M. Raghu, and S. Bengio, Insights on representational similarity in neural networks with canonical correlation, in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Vol. 31 (Curran Associates, Montréal, 2018), pp. 5727–5736.

[6] S. Kornblith *et al.*, Similarity of neural network representations revisited, in *Proceedings of the 36th International Conference on Machine Learning*, edited by K. Chaudhuri and R. Salakhutdinov, Proceedings of Machine Learning Research, Vol. 97 (PMLR, Long Beach, 2019), pp. 3519–3529.

[7] Y. Li *et al.*, Convergent learning: Do different neural networks learn the same representations? in *Proceedings of the 1st International Workshop on Feature Extraction: Modern Questions and Challenges at NIPS 2015*, edited by D. Storcheus, A. Rostamizadeh, and S. Kumar, Proceedings of Machine Learning Research, Vol. 44 (PMLR, Montreal, 2015), pp. 196–212.

[8] J. V. Haxby *et al.*, Distributed and overlapping representations of faces and objects in ventral temporal cortex, Science **293**, 2425 (2001) 2430.

[9] N. Kriegeskorte *et al.*, Matching categorical object representations in inferior temporal cortex of man and monkey, Neuron **60**, 1126 (2008).

[10] D. L. Yamins *et al.*, Hierarchical modular optimization of convolutional networks achieves representations similar to macaque IT and human ventral stream, in *Advances in Neural Information Processing Systems*, edited by C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Vol. 26 (Curran Associates, Lake Tahoe, 2013), pp. 3318–3126.

[11] D. Sussillo *et al.*, A neural network that finds a naturalistic solution for the production of muscle activity, Nat. Neurosci. **18**, 1025 (2015).

[12] F. Du *et al.*, Towards a simplified model of primary visual cortex, bioRxiv (2024), doi:10.1101/2024.06.30.601394.

[13] I. Sucholutsky *et al.*, Getting aligned on representational alignment, arXiv:2310.13018.

[14] P. Robert and Y. Escoufier, A unifying tool for linear multivariate statistical methods: The RV coefficient, J. R. Stat. Soc. C: Appl. Stat. **25**, 257 (1976).

[15] A. Cichocki, R. Zdunek, and S.-I. Amari, Nonnegative matrix and tensor factorization, IEEE Signal Proc. Mag. **25**, 142 (2007).

[16] A. Cichocki and S.-I. Amari, Families of alpha-, beta- and gamma-divergences: Flexible and robust measures of similarities, Entropy **12**, 1532 (2010).

[17] C. Cortes, M. Mohri, and A. Rostamizadeh, Algorithms for learning kernels based on centered alignment, J. Mach. Learn. Res. **13**, 795 (2012).

[18] S. Tang *et al.*, Similarity of neural networks with gradients, arXiv:2003.11498.

[19] T. Gärtner, P. Flach, and S. Wrobel, On graph kernels: Hardness results and efficient alternatives, in *Proceedings of Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24–27* (Springer, 2003), pp. 129–143.

[20] K. M. Borgwardt and H.-P. Kriegel, Shortest-path kernels on graphs, in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, edited by J. Han, B. Wah, V. Raghaven, X. Wu, R. Rastogi (IEEE, Houston, 2005), pp. 78–81.

[21] N. Shervashidze *et al.*, Weisfeiler-Lehman graph kernels, J. Mach. Learn. Res. **12** (2011).

[22] A. Gretton *et al.*, A kernel two-sample test, J. Mach. Learn. Res. **13**, 723 (2012).

[23] N. M. Kriege, F. D. Johansson, and C. Morris, A survey on graph kernels, Appl. Network Sci. **5**, 6 (2020).

[24] H. Hotelling, Relations between two sets of variates, Biometrika **28**, 321 (1936).

[25] G. H. Golub and H. Zha, The canonical correlations of matrix pairs and their numerical computation, in *Linear Algebra for Signal Processing*, IMA Volumes in Mathematics and Its Applications, edited by A. Bojanczyk and G. Cybenk, Vol. 69 (Springer, New York, 1995), pp. 27–49.

[26] A. Andreella *et al.*, Procrustes-based distances for exploring between-matrices similarity, Stat. Methods Appl. **32**, 867 (2023).

[27] P. H. Schönemann, A generalized solution of the orthogonal procrustes problem, Psychometrika **31**, 1 (1966).

[28] A. H. Williams *et al.*, Generalized shape metrics on neural representations, in *Advances in Neural Information Processing Systems*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, Vol. 34 (Curran Associates, 2021), pp. 4738–4750.

[29] F. Ding, J.-S. Denain, and J. Steinhardt, Grounding representation similarity through statistical testing, in *Advances in Neural Information Processing Systems*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, Vol. 34 (Curran Associates, 2021), pp. 1556–1568.

[30] N. Cloos *et al.*, Differentiable optimization of similarity scores between models and brains, arXiv:2407.07059.

[31] A. Barachant *et al.*, Multiclass brain-computer interface classification by Riemannian geometry, IEEE Trans. Biomed. Eng. **59**, 920 (2011).

[32] C. Ju and C. Guan, Tensor-cspnet: A novel geometric deep learning framework for motor imagery classification, IEEE Trans. Neural Netw. Learning Syst. **34**, 10955 (2022).

[33] R. Gutzen, S. Grün, and M. Denker, Evaluating the statistical similarity of neural network activity and connectivity via eigenvector angles, BioSystems **223**, 104813 (2023).

[34] L. N. Trefethen and D. Bau, *Numerical Linear Algebra, Other Titles in Applied Mathematics*, Vol. 181 (SIAM, Philadelphia, USA, 1997), pp. 25–30.

[35] C. Jordan, Essai sur la géométrie à *n* dimensions, Bull. Société math. France **3**, 103 (1875).

[36] P. Zhu and A. V. Knyazev, Angles between subspaces and their tangents, J. Num. Math. **21**, 325 (2013).

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PRXLife.3.023005 for additional derivations, comparison with other well-known similarity metrics, and the application of SAS to nonsquare synthetic matrices.

[38] N. Halko, P. G. Martinsson, and J. A. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, SIAM Rev. **53**, 217 (2011).

[39] G. H. Golub, A. Hoffman, and G. W. Stewart, A generalization of the Eckart-Young-Mirsky matrix approximation theorem, Linear Algebra Appl. **88–89**, 317 (1987).

[40] A. H. Williams *et al.*, Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping, Neuron **105**, 246 (2020).

[41] M. Potters and J.-P. Bouchaud, *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists* (Cambridge University Press, Cambridge, 2020).

[42] T. Tao, *Topics in Random Matrix Theory*, Graduate Studies in Mathematics, Vol. 132 (American Mathematical Society, Rhode Island, 2012).

[43] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, J. Math. Phys. **6**, 440 (1965).

[44] P. Erdős and A. Rényi, On random graphs, Publ. Math. **6**, 290 (1959).

[45] J. Park and M. E. J. Newman, Statistical mechanics of networks, Phys. Rev. E **70**, 066117 (2004).

[46] C. Cooper and A. Frieze, The size of the largest strongly connected component of a random digraph with a given degree sequence, Comb. Probab. Comput. 13, 319 (2004).

[47] D. J. Watts and S. H. Strogatz, Collective dynamics of small-world networks, Nature (London) **393**, 440 (1998).

[48] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).

[49] X. Chen *et al.*, 1024-channel electrophysiological recordings in macaque V1 and V4 during resting state, Sci. Data **9**, 77 (2022).

[50] H. Supèr and P. R. Roelfsema, Chronic multiunit recordings in behaving animals: Advantages and limitations, in *Progress in Brain Research*, Vol. 147 (Elsevier, 2005), pp. 263–282.

[51] A. Morales-Gregorio *et al.*, Neural manifolds in V1 change with top-down signals from V4 targeting the foveal region, Cell Rep. **43**, 114371 (2024).

[52] V. A. Marchenko and L. A. Pastur, Distribution of eigenvalues for some sets of random matrices, Math. USSR-Sb. **1**, 457 (1967).

[53] J. Cohen, *Statistical Power Analysis for the The Behavioral Sciences* (L. Erlbaum Associates, 1988).

[54] Student [William Sealy Gosset], The probable error of a mean, Biometrika **6**, 1 (1908).

[55] B. L. Welch, The generalization of Student's problem when several different population varlances are involved, Biometrika **34**, 28 (1947).

[56] T. Cai, J. Fan, and T. Jiang, Distributions of angles in random packing on spheres, J. Mach. Learn. Res. **14**, 1837 (2013).

[57] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, Vol. 3 (Pergamon Press, Oxford, 1958), pp. 129–123.

[58] P. Deuflhard and A. Hohmann, *Numerische Mathematik 1: Eine algorithmisch orientierte Einführung* (Walter de Gruyter, Berlin, 2008), pp. 152–158.

[59] R. B. Tootell *et al.*, Deoxyglucose analysis of retinotopic organization in primate striate cortex, Science **218**, 902 (1982).

[60] D. H. Hubel and T. N. Wiesel, Receptive fields of single neurones in the cat's striate cortex, J. Physiol. **148**, 574 (1959).

[61] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, Spatial frequency selectivity of cells in macaque visual cortex, Vision Res. **22**, 545 (1982).

[62] M. Ostrow *et al.*, Beyond geometry: Comparing the temporal structure of computation in neural circuits with dynamical similarity analysis, in *Advances in Neural Information Processing Systems*, edited by A. Oh, T. Naumann, A. Globerson, K.

Saenko, M. Hardt, and S. Levine, Vol. 36 (Curran Associates, New Orleans, 2023), pp. 33824–33837.
[63] S. L. Brunton *et al.*, Modern Koopman theory for dynamical systems, arXiv:2102.12086.
[64] Q. Guilhot *et al.*, Dynamical similarity analysis uniquely captures how computations develop in RNNs, arXiv:2410.24070.
[65] K. Scharnhorst, Angles in complex vector spaces, Acta Appl. Math. **69**, 95 (2001).
[66] S. Kritchman and B. Nadler, Determining the number of components in a factor model from limited noisy data, Chemom. Intell. Lab. Syst. **94**, 19 (2008).
[67] S. Kritchman and B. Nadler, Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory, IEEE Trans. Signal Proc. **57**, 3930 (2009).
[68] J. Veraart, E. Fieremans, and D. S. Novikov, Diffusion MRI noise mapping using random matrix theory, Magn. Reson. Med. **76**, 1582 (2016).
[69] S. Safavi, N. K. Logothetis, and M. Besserve, From univariate to multivariate coupling between continuous signals and point processes: A mathematical framework, Neural Comput. **33**, 1751 (2021).
[70] S. Safavi *et al.*, Uncovering the organization of neural circuits with generalized phase locking analysis, PLOS Comput. Biol. **19**, e1010983 (2023).
[71] J. Bun, J.-P. Bouchaud, and M. Potters, Overlaps between eigenvectors of correlated random matrices, Phys. Rev. E **98**, 052145 (2018).
[72] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica **37**, 424 (1969).
[73] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. **85**, 461 (2000).
[74] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms* (MIT Press, Cambridge, MA, 2017).
[75] https://github.com/INM-6/SAS.
[76] https://doi.org/10.5281/zenodo.10680478.
[77] R. Gutzen *et al.*, Reproducible neural network simulations: Statistical methods for model validation on the level of network activity data, Front. Neuroinformatics **12**, 90 (2018).
[78] https://doi.org/10.5281/zenodo.10680810.
[79] https://github.com/INM-6/NetworkUnit.