

PAPER • OPEN ACCESS

Clustering matrices through optimal permutations

To cite this article: Flaviano Morone 2022 *J. Phys. Complex.* 3 035007

View the [article online](#) for updates and enhancements.

You may also like

- [Coincidence complex networks](#)
Luciano da Fontoura Costa
- [A random matrix perspective of cultural structure: groups or redundancies?](#)
Alexandru-Ionu Bbeanu
- [A microfluidic device for the continuous culture and analysis of *Caenorhabditis elegans* in a toxic aqueous environment](#)
Jaehoon Jung, Masahiro Nakajima, Hirotaka Tajima et al.

OPEN ACCESS

PAPER



Clustering matrices through optimal permutations

RECEIVED

25 May 2022

REVISED

10 August 2022

ACCEPTED FOR PUBLICATION

24 August 2022

PUBLISHED

15 September 2022

Flaviano Morone*

Department of Physics, New York University, New York, NY 10003, United States of America

* Author to whom any correspondence should be addressed.

E-mail: fm2452@nyu.edu**Keywords:** clusterings, optimization, permutations, doubly-stochastic matrices, quadratic assignment

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Abstract**

Matrices are two-dimensional data structures allowing one to conceptually organize information. For example, adjacency matrices are useful to store the links of a network; correlation matrices are simple ways to arrange gene co-expression data or correlations of neuronal activities. Clustering matrix entries into geometric patterns that are easy to interpret helps us to understand and explain the functional and structural organization of the system components described by matrix entries. Here we introduce a theoretical framework to cluster a matrix into a desired pattern by performing a similarity transformation obtained by solving an optimization problem named optimal permutation problem. On the numerical side, we present an efficient clustering algorithm that can be applied to any type of matrix, including non-normal and singular matrices.

We apply our algorithm to the neuronal correlation matrix and the synaptic adjacency matrix of the *Caenorhabditis elegans* nervous system by performing different types of clustering, using block-diagonal, nested, banded, and triangular patterns. Some of these clustering patterns show their biological significance in that they separate matrix entries into groups that match the experimentally known classification of *C. elegans* neurons into four broad categories made up of interneurons, motor, sensory, and polymodal neurons.

1. Introduction

If a block of data is endowed with a binary relational structure, then it is usually possible to exploit a matrix to store it [1]. For example, two-point correlations can be stored into a covariance matrix [2, 3]; the links between the nodes of a graph into an adjacency matrix [4]; a QR code into a matrix bar code [5]. However, the resulting matrix often appears incomprehensible to our visual inspection and does not manifest the global information it contains, nor the details, to our eyes.

Clustering a matrix into a desired geometric pattern [6] allows us to see clearly and understand the whole information and the details in one fell swoop. And the more clearly we see the whole in a single glance, the better we perceive its analogies with other related objects, and therefore the more opportunities we have to grasp the possible generalizations. Thus, clustering data sharpens up our scientific understanding of nature, simultaneously giving our eyes a soothing aesthetic satisfaction.

Here, we introduce a theoretical framework to find an optimal permutation of the rows and columns of a square matrix that brings such matrix as close as possible to a desired clustered form. The problem amounts to minimize an objective function of a permutation matrix and for this reason we name it the optimal permutation problem (OPP). We formulate the OPP in the language of statistical mechanics, whereby we can find its solution by computing the partition function through the saddle-point method.

To set the stage of our discussion, we introduce the OPP in a pragmatic way by considering a matrix B , that we call *input matrix* and show in figure 1(a), describing the correlation of neuronal activities of $N = 33$ neurons of the nematode *C. elegans* measured for three locomotory tasks of the animal (forward, backward, and turn) in reference [3]. The choice of this particular dataset is useful to get an instrumental view of the OPP and how it relates to real-world data, although we could formulate our mathematical theory completely

in abstracto. Thus, we emphasize that the optimal permutation theory and the algorithm we present here can be applied to any square matrix of the most general form.

The entries B_{ij} of the correlation matrix in figure 1(a) take values in the interval $B_{ij} \in [-1, 1]$, where the extreme value $B_{ij} = 1$ occurs for neurons i and j which are both active during locomotion. The other extreme value $B_{ij} = -1$, instead, occurs whenever neuron i is active while neuron j is not, and vice versa. The existence of positive and negative correlations implies the existence of at least two groups of neurons such that all neurons in one group are positively correlated with each other and negatively correlated with neurons in the other group. The presence of two groups of neurons can be traced back in the twofold nature of the locomotion behavior comprising: forward movement mediated by one group of neurons, and backward movement mediated by another one (turns/reversals are accounted for by a third group of neurons, as we shall discuss at the end of this section).

To identify the two groups, we consider a matrix A , called *filter matrix*, with a two-blocks shape, as seen in figure 1(b). Precisely, $A_{ij} = 1$ if $i, j \in [1, N/2] \times [1, N/2] \cup [N/2 + 1, N] \times [N/2 + 1, N]$, and $A_{ij} = 0$ otherwise. The role of the filter matrix A is to conceptualize visually how the input matrix B ought to be clustered into two blocks. In other words, we use the filter matrix A to guide the clustering process in order to bring the input matrix B into a clustered form B' as 'similar' as possible to the shape of A . Mathematically, this can be achieved by means of an objective function $E(P)$ defined as

$$E(P) = \|PA - BP\|^2 = \|A\|^2 + \|B\|^2 - 2 \text{Tr}(B^t P A P^t), \quad (1)$$

where $\|A\|^2 = \text{Tr}(A^t A)$ is the Frobenius norm and P is a permutation matrix, whose entries $P_{ij} \in \{0, 1\}$ satisfy the constraints $\sum_i P_{ij} = \sum_j P_{ij} = 1$. The objective function in equation (1) appeared for the first time in the formulation of the quadratic assignment problem (QAP) [7], which is one of the most important problems in the field of combinatorial optimization [8]. The QAP was initially introduced in economics to find the optimal assignment (=optimal permutation P in our language) of N facilities to N locations, given the matrix of distances between pair of facilities (=filter matrix A) and a weight matrix quantifying the amount of goods flowing between firms (=input matrix B).

To better explain the meaning of the objective function in equation (1), let us suppose that we could find a permutation matrix P_* such that $E(P_*) = 0$. This means that matrix A is permutation similar to matrix B via the transformation $A = P_*^t B P_*$. That is, A itself is an exact clustered form of the matrix B . If the equation $E(P) = 0$ has no solution, then A is not permutation similar to B ; hence A is not itself an exact clustering of B . In such a state of affairs the best we can do is to look for a permutation matrix P_* that still minimizes the disagreement between A and B , albeit not zeroing the objective function $E(P)$, i.e. $\min_P E(P) = E(P_*) > 0$, as seen in figure 1(d). Thus, we define P_* to be the solution of the following optimization problem

$$P_* = \arg \min_P E(P). \quad (2)$$

We call P_* the **optimal permutation** of the input matrix B (given the filter matrix A) and we show it in figure 1(e). Once we obtain P_* we can proceed to cluster B by performing a similarity transformation to bring B into its clustered form B' as

$$B' = P_*^t B P_*. \quad (3)$$

The result is shown in figure 1(f). The two-blocks clustering (left panel in figure 1(f)) identifies two clusters separating two groups of neurons: one group contains neurons driving backward locomotion, and the other one contains those regulating forward locomotion.

By using the three-blocks filter shown in figure 1(c) we obtain a clustering of the correlation matrix B into three clusters: two of them are each a subset of the backward and forward locomotion groups found previously. The third cluster occupies the middle block in the right panel of figure 1(f). Neurons belonging to this cluster are classified by the Wormatlas database [9] as: ring interneurons (RIVL/RIVR) regulating reversals and deep omega-shaped turns; motor neurons (SMDVR/SMDVL, RMEV) defining the amplitude of omega turns; labial neurons (OLQDR/OLQVL) regulating nose oscillations in local search behavior; and a high-threshold mechanosensor (ALA) responding to harsh-touch mechanical stimuli. We term 'turn' the third block in the clustered correlation matrix shown in figure 1(f).

Having formulated the OPP, we move now to explain the analytical and numerical methods to solve it and then discuss several applications for clustering the *C. elegans* whole brain's connectome.

2. How to solve the optimal permutation problem

In order to determine a solution to the OPP given in equation (2) we use statistical mechanics methods [10]. The quantity which plays the fundamental role in the resolution of the OPP is the partition function $Z(\beta)$,

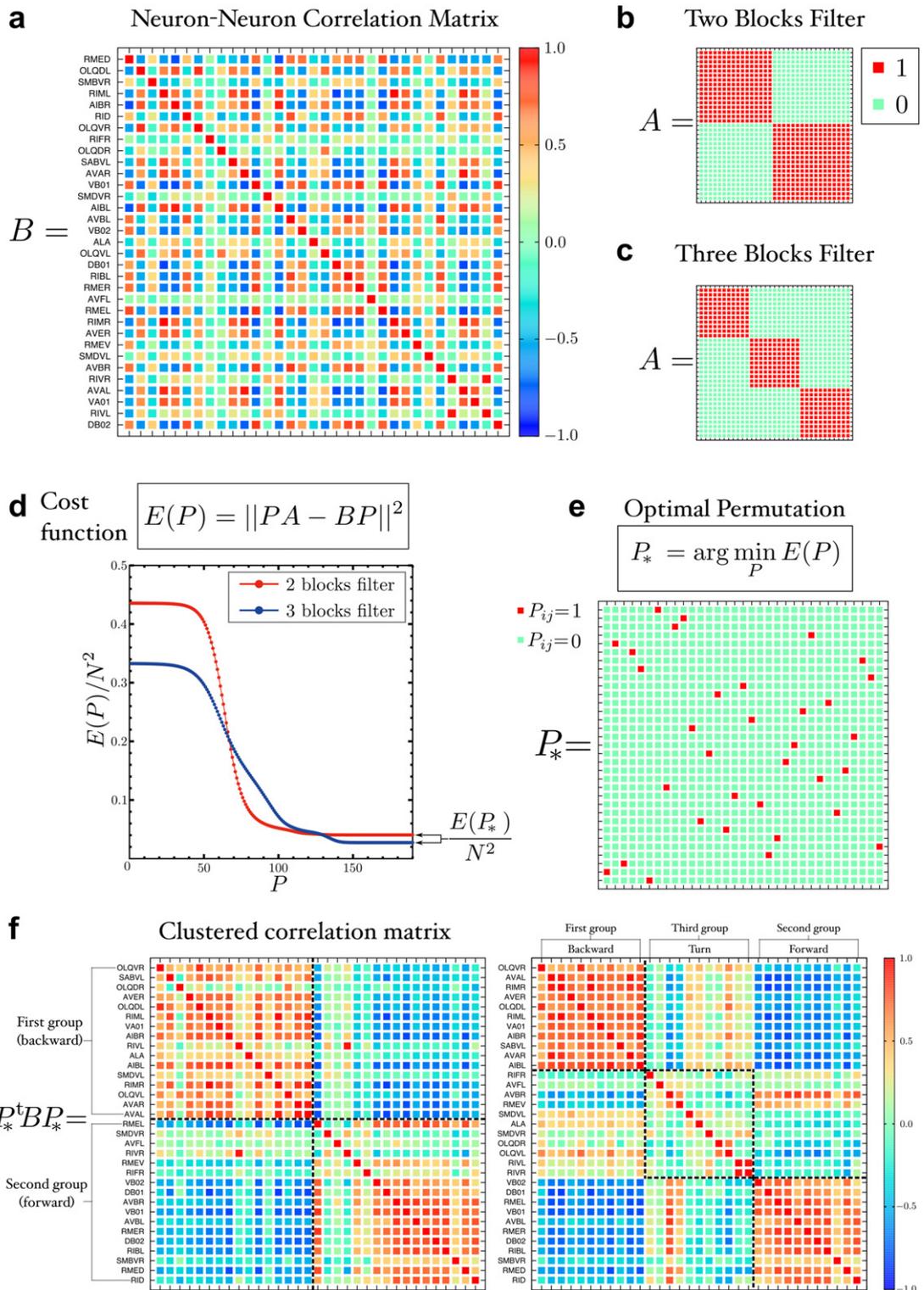


Figure 1. Explanation of the OPP. (a) Correlation matrix of the neuronal activity of the *C. elegans*. Each entry B_{ij} is the correlation coefficient between the time series x_i^t and x_j^t (data are from reference [3]). (b) Two blocks filter matrix A to be applied to the input matrix B to perform the clustering of B into two blocks, each one made up of neurons maximally correlated among each other. The two blocks are arranged along the main diagonal. (c) Three blocks filter which, similarly to the filter in (b), produces a clustering of B into three clusters. (d) Minimization of the cost function $E(P)$ defined in equation (1) for two and three-blocks filters. The horizontal axis indexes the permutation matrix P found at any given iteration step t of the algorithm. That is, at iteration $t = 1$ we find the permutation matrix $P = P_1$ and we compute the corresponding energy $E(P_1)/N^2$; at the next iteration $t = 2$ we find P_2 and we calculate $E(P_2)/N^2$; and so on and so forth until convergence takes place at P_* corresponding to the minimum of the energy $E(P_*)/N^2$ (see appendix B for the precise definition of the convergence criterion). (e) The optimal permutation matrix P_* that solves the OPP defined in equation (2) for the correlation matrix B shown in (a) and the two-blocks filter A in (b). The permutation P_* is the one that minimizes the cost function in (d) (red curve), i.e., $P_* : \min_P E(P) = E(P_*)$. (f) Clustered correlation matrix $B' = P_*^t B P_*^t$ obtained by solving the OPP with a two-blocks filter (left panel) and a three-blocks filter (right panel).

defined as

$$Z(\beta) = \sum_P e^{-\beta E(P)}, \quad (4)$$

where the sum is over all permutation matrices P belonging to the symmetric group on N elements. The statistical physics interpretation of the problem thus follows. The parameter β in equation (4) represents the inverse of the ‘temperature’ of the system; the objective function $E(P)$ defined in equation (1) becomes the ‘energy’ function. The global minimum of the energy function corresponds to the ‘ground state energy’ of the system. Since a physical system goes into its ground state only at zero temperature (by the third law of thermodynamics), then the exact solution to the OPP corresponds to the zero temperature limit of the partition function in equation (4), i.e.

$$\lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \log Z(\beta) = \min_P E(P) = E(P_*). \quad (5)$$

The concept of temperature is somewhat artificial and it is introduced in order to recast the problem in the language of the canonical ensemble of statistical mechanics. In the canonical ensemble we calculate the partition function by summing over the set of ‘microstates’ P with N and β fixed, but we allow the energy to fluctuate. It could be possible, in principle, to solve the problem in the framework of the microcanonical ensemble. In this ensemble we sum over the set of microstates with N and E fixed, while β is a derived quantity. The main reason to work in the canonical ensemble is that we do not know how to fix the ground-state energy *a priori*: it is the unknown we wish to find. On the contrary, we can fix the temperature, because we know that the system attains its ground state energy exactly at zero temperature; thus we know we have to send $\beta \rightarrow \infty$ at the end of the calculation of the partition function.

We like to make an important observation regarding the ground state P_* and the corresponding global minimum $E(P_*)$. That is: the energy function has, in general, many degenerate global minima. To prove this claim, let us suppose that there exists a non trivial permutation matrix Q that commutes with the input matrix B , i.e. $[B, Q] = BQ - QB = 0$. By multiplying P_* and Q we obtain another permutation matrix, denoted as $R = QP_*$. The energy $E(R)$ of the new state R is given by (neglecting terms independent from R)

$$E(R) = -2 \operatorname{Tr}(B^t R A R^t) = -2 \operatorname{Tr}(B^t Q P_* A P_*^t Q^t) = -2 \operatorname{Tr}(B^t P_* A P_*^t) = E(P_*), \quad (6)$$

meaning that state R has the same energy of the ground state P_* ; hence the ground state is degenerate. In general, the degeneracy of the ground state is at least equal to the order of the automorphism group of B times the order of the automorphism group of A , i.e. $|\operatorname{Aut}(B)| \times |\operatorname{Aut}(A)|$.

Next we derive an integral representation of the partition function that can be evaluated by the steepest descent method in the limit $\beta \rightarrow \infty$.

2.1. Integral representation of the partition function

Since the partition function in equation (4) can be easily calculated when all P_{ij} appear linearly in the argument of the exponential, a good idea is to write the quadratic term, which in the energy connects two variables P_{ij} and $P_{k\ell}$, as an integral over disconnected terms [10]. In order to achieve this result, we insert the δ -function

$$\delta(X_{ij} - P_{ij}) = \frac{1}{2\pi i} \int dY_{ij} e^{-Y_{ij}(X_{ij} - P_{ij})}, \quad (7)$$

where the integration over Y_{ij} runs along the imaginary axis, into the representation of the partition function

$$Z(\beta) \propto \sum_P \int \prod_{ij} dX_{ij} \int \prod_{ij} Y_{ij} e^{-\beta E(X) + \sum_{ij} Y_{ij}(P_{ij} - X_{ij})}, \quad (8)$$

where $E(X)$ is the energy function defined by

$$E(X) = -\frac{1}{2} \operatorname{Tr}(B^t X A X^t), \quad (9)$$

and the prefactor $1/2$ has been introduced only for convenience. To proceed further in the calculation, we enforce the constraint on the columns sums of P (and thus on the columns sums of X) by inserting N δ -functions

$$\delta \left(\sum_i X_{ij} - 1 \right) = \frac{1}{2\pi i} \int dz_j e^{-z_j \left(\sum_i X_{ij} - 1 \right)}, \tag{10}$$

one for each $j = 1, \dots, N$ into the partition function, thus obtaining

$$Z(\beta) \propto \sum_P' \int \prod_{ij} dX_{ij} \int \prod_{ij} dY_{ij} \int \prod_j dz_j e^{-\beta E(X) + \sum_{ij} Y_{ij}(P_{ij} - X_{ij}) - \sum_j z_j \left(\sum_i X_{ij} - 1 \right)}. \tag{11}$$

The prime (') written above the sum symbol in equation (11) has the following explanation. By enforcing the constraints on the columns sums through the Lagrange multipliers z_j , the sum over P now runs over the set of matrices with entries $P_{ij} \in \{0, 1\}$ subject only to the constraint that rows sum to one. The Lagrange multiplier z_j is called ‘chemical potential’ in the language of statistical mechanics. By fixing the Lagrange multipliers we allow the columns sums of P to vary; similarly, in statistical mechanics, by fixing the chemical potential we allow the number of particles in the system to vary (when the particles’ number can vary we are technically using the grand canonical ensemble).

That being said, summation over the variables P_{ij} is now straightforward and we find

$$\sum_P e^{\sum_{ij} Y_{ij} P_{ij}} = \prod_i \sum_j e^{Y_{ij}}. \tag{12}$$

Let us designate by $F(X, Y, z)$ the free-energy function defined by

$$F(X, Y, z) = E(X) + \frac{1}{\beta} \text{Tr}(XY^t) - \frac{1}{\beta} \sum_i \log \sum_j e^{Y_{ij}} + \frac{1}{\beta} \sum_j z_j \left(\sum_i X_{ij} - 1 \right), \tag{13}$$

whereby we can write the partition function as

$$Z(\beta) \propto \int \prod_{ij} dX_{ij} dY_{ij} \prod_j dz_j e^{-\beta F(X, Y, z)}, \tag{14}$$

which can be evaluated by the saddle-point method in the limit $\beta \rightarrow \infty$.

2.2. Saddle-point equations

The saddle point equations are obtained by differentiating F with respect to X_{ij} , Y_{ij} , and z_j :

$$\begin{aligned} \frac{\partial F}{\partial X_{ij}} &= \frac{\partial E}{\partial X_{ij}} + \frac{1}{\beta} Y_{ij} + \frac{1}{\beta} z_j = 0, \\ \frac{\partial F}{\partial Y_{ij}} &= \frac{1}{\beta} X_{ij} - \frac{1}{\beta} \frac{e^{Y_{ij}}}{\sum_k e^{Y_{ik}}} = 0, \\ \frac{\partial F}{\partial z_j} &= \frac{1}{\beta} \left(\sum_i X_{ij} - 1 \right) = 0. \end{aligned} \tag{15}$$

The solution to the saddle point equation (15) is given by:

$$\begin{aligned} Y_{ij} &= \frac{\beta}{2} (BXA^T + B^T XA)_{ij} - z_j, \\ X_{ij} &= \frac{e^{Y_{ij}}}{\sum_k e^{Y_{ik}}}, \\ 1 &= \sum_i X_{ij}. \end{aligned} \tag{16}$$

From the second of equation (16) we see that the solution X_{ij} satisfies automatically the condition of having rows summing to one, as it should. Opposed to this, the constraints on the columns sums, expressed by the third of equation (16), are taken into account by the Lagrange multipliers z_j . It is worth noticing that the equation for X_{ij} is invariant under global translations of z_j of the form

$$z_j \rightarrow z_j + c, \quad \forall j, \quad (17)$$

for arbitrary values of c . This symmetry is not unexpected, though. Indeed, it can be traced back to the fact that out of the $2N$ constraints on the rows and columns sums of P (and thus of X) only $2N - 1$ of them are linearly independent, since the sum of all entries must be equal to N , i.e. $\sum_{ij} P_{ij} = \sum_{ij} X_{ij} = N$. This translational symmetry can be eliminated, for example, by choosing c in such a way that $\sum_j z_j = 0$. However, we shall derive next a set of simplified equations to compute the Lagrange multipliers that does not require to fix c explicitly. To this end, we introduce the matrix $W(X)$ defined by

$$W_{ij}(X) = \exp \left[\frac{\beta}{2} (BXA^t + B^tXA)_{ij} \right], \quad (18)$$

along with two vectors: a ‘right’ vector with components V_j given by

$$V_j = e^{-z_j}, \quad (19)$$

and a ‘left’ vector with components U_i given by

$$U_i = \frac{1}{\sum_j W_{ij}(X) V_j}, \quad (20)$$

whereby we can write the saddle point equations as a set of fixed point equations involving only X :

$$X_{ij} = U_i(X) \exp \left[\frac{\beta}{2} (BXA^t + B^tXA)_{ij} \right] V_j(X). \quad (21)$$

Vector V can be calculated consistently with U by solving the following equations

$$V_j = \frac{1}{\sum_i U_i W_{ij}(X)}. \quad (22)$$

It is interesting to notice that equations (20) and (22) coincide with the Sinkhorn–Knopp equations [11, 12] for rescaling to one the rows and columns sums of a matrix with strictly positive entries (such is indeed the matrix $W_{ij}(X)$). We also note, en passant, that equations similar to (20) and (22) have been introduced in reference [13] to simultaneously measure the economic competitiveness of countries ($=U_i^{-1}$) and the sophistication ($=V_j$) of products they export in world trade.

Eventually, to find the optimal permutation matrix P_* we have to take the zero temperature limit by sending $\beta \rightarrow \infty$: in this limit the solution matrix $X = X(\beta)$ is projected onto one of the $N!$ vertices of the hypercube $[0, 1]^{N^2}$. For large but finite β , the solution $X(\beta)$ is a doubly-stochastic matrix, which approximates the optimal permutation matrix P_* . Equation (21) represents our main result, along with the algorithm to solve it that is described in detail in appendix B.

In the rest of this section we shall try to elucidate how the saddle point solution $X(\beta)$ is related to the solution of the original problem P_* . The crux is that equation (15) has, in general, many solutions at low temperature. How are these solutions related to the zero temperature optimal solution of the original problem? The existence of many solutions is invariably related to the existence of many local minima of the free-energy function at low temperature. Specifically, the free-energy given by equation (13) is convex at sufficiently high temperatures (i.e. for $\beta \ll 1$), but will typically have many local minima at low temperature. This can be seen by considering the free-energy projected onto the space obeying the global constraints (see appendix A for the calculation of the projected free-energy), that reads

$$F(X) = -\frac{1}{2} \text{Tr}(B^tXAX^t) + \frac{1}{\beta} \sum_{ij} X_{ij} \log X_{ij}. \quad (23)$$

The stability of the saddle point solution X_* is determined by the Hessian $H_{ij;kl}$ which is given by

$$H_{ij;kl} = \left. \frac{\partial^2 F}{\partial X_{ij} \partial X_{kl}} \right|_{X_*} = -B_{ik}A_{jl} + \frac{T}{(X_*)_{ij}} \delta_{ik} \delta_{jl}, \quad (24)$$

where $T = \beta^{-1}$ is the temperature. The solution X_* is stable if the Hessian is positive definite. This is certainly true at sufficiently high temperature (i.e. for $T \gg 1$), because the second term dominates. However, by decreasing the temperature the first term dominates the Hessian and the solution X_* becomes unstable at a finite temperature $T_c > 0$ if the term $B_{ik}A_{jl}$ has a positive eigenvalue in its spectrum. Below T_c , the solution X_* bifurcates and therefore the free-energy will have many local minima. If these bifurcations occur

in the limit of infinite system size $N \rightarrow \infty$, the system undergoes a series of phase transitions (in the statistical mechanics language) as the temperature is lowered down to zero. These bifurcations have a major impact on the performance of any optimization algorithm which attempts to track the solution (and thus the minimum of the energy) as the temperature decreases, since there is no guarantee that the minimum at temperature T can be matched to the minimum of the original problem at $T = 0$ [14, 15]. Our algorithm is no exception. Although a complete analysis on convergence and bounds of our algorithm is beyond the scope of this paper, the numerical findings discussed in the next section are encouraging enough to support our belief that it can be a good heuristic algorithm for obtaining nearly optimal solutions to difficult clustering problems.

Next we move to apply our algorithm to the *C. elegans* connectome by performing three types of clustering to its adjacency matrix.

3. Clustering the *C. elegans* connectome

We consider the neuronal network of the hermaphrodite *C. elegans* comprising $N = 253$ neurons interconnected via gap junctions (we consider only the giant component of this network). We use the most up-to-date connectome of gap-junctions from reference [16]. We represent the synaptic connectivity structure via a binary adjacency matrix B , with $B_{ij} = 1$ if neuron i connects (i.e. forms gap-junctions) to j , and $B_{ij} = 0$ otherwise, as shown in figure 2(a) ($\sum_{ij} B_{ij} = 2M = 1028$, so that the mean degree is $\langle k \rangle = 2M/N \sim 4$). Gap-junctions are undirected links, hence B is a symmetric matrix. We emphasize that the application of our framework is not limited to symmetric matrices and can be equally applied to asymmetric adjacency matrices representing directed chemical synapses.

We perform a clustering experiment by using a filter A whose shape is shown in figure 2(b). We call it: ‘nestedness filter’. This nomenclature is motivated by ecological studies of species abundance showing nested patterns in the community structure of mammals [17] and plant-pollinator ecosystems [18, 19]. Nestedness is also found in interbank, communication, and socio-economic networks [13, 20, 21]. Remarkably, it has been shown recently that behavioral control in *C. elegans* occurs via a nested neuronal dynamics across motor circuits [22]. A connectivity structure which is nested implies the existence of two types of nodes, that are called ‘generalists’ and ‘specialists’. Generalists are ubiquitous species with a large number of links to other species that are quickly reachable by the other nodes; specialists are rare species with a small number of connections occupying peripheral locations of the network and having higher extinction risk [23].

The entries of A are defined by:

$$A_{ij} = \begin{cases} A_{ij} = 1 & \text{if } j \leq f(i; p) = N - (i - 1)^p(N - 1)^{1-p} \\ A_{ij} = 0 & \text{ow} \end{cases}, \quad (25)$$

where p is the nestedness exponent controlling the curvature of the function $f(i; p)$ separating the filled and empty parts of the matrix A , as seen in figure 2(b). By solving the OPP we obtain the optimal permutation matrix P_* shown in figure 2(c), whereby we cluster the adjacency matrix B via the similarity transformation $B' = P_*^t B P_*$, as shown in figure 2(d). In order to measure the degree of nestedness of the connectome we introduce the quantity $\phi(p)$, defined as the fraction of entries of B' comprised in the nested region $j \leq f(i; p)$ by the following formula

$$\phi(p) = \frac{\sum_{i=1}^N \sum_{j=1}^{f(i;p)} (P_*^t B P_*)_{ij}}{\sum_{i=1}^N \sum_{j=1}^N (P_*^t B P_*)_{ij}} = \frac{1}{2M} \sum_{i=1}^N \sum_{j=1}^{f(i;p)} (P_*^t B P_*)_{ij}. \quad (26)$$

We call $\phi(p)$ the ‘packing fraction’ of the connectome. The profile of $\phi(p)$ as a function of p is shown in figure 2(e), comparing the *C. elegans* connectome to a randomized connectome where neurons have the same degree as in the original connectome, but links are rewired at random through the configurational model [4]. Figure 2(e) shows that the *C. elegans* connectome is 10% more packed than its random counterpart almost for every value of p in the range $(0, 1]$. Lastly, in figure 2(f) we separate the neurons into two groups as follows: generalists for $i = 1, \dots, N/2$, and specialists for $i = N/2 + 1, \dots, N$. We find that: 3/4 of interneurons are generalists and only 1/4 are specialists; motor neurons are split nearly half and half between generalists and specialists; while 2/3 of sensory and polymodal neurons are specialists and 1/3 are generalists (broad functional categories of neurons—(1) interneurons, (2) motor neurons, (3) sensory neurons, and (4) polymodal neurons—are compiled and provided at <http://wormatlas.org/hermaphrodite/nervous/Neuroframeset.html>, chapter 2.2 [9]).

Last but not least, we present three more clustering of the *C. elegans* connectome, performed via three filters: the bandwidth filter [24], showed in figure 3(a); the triangular filter in figure 3(b); and the square

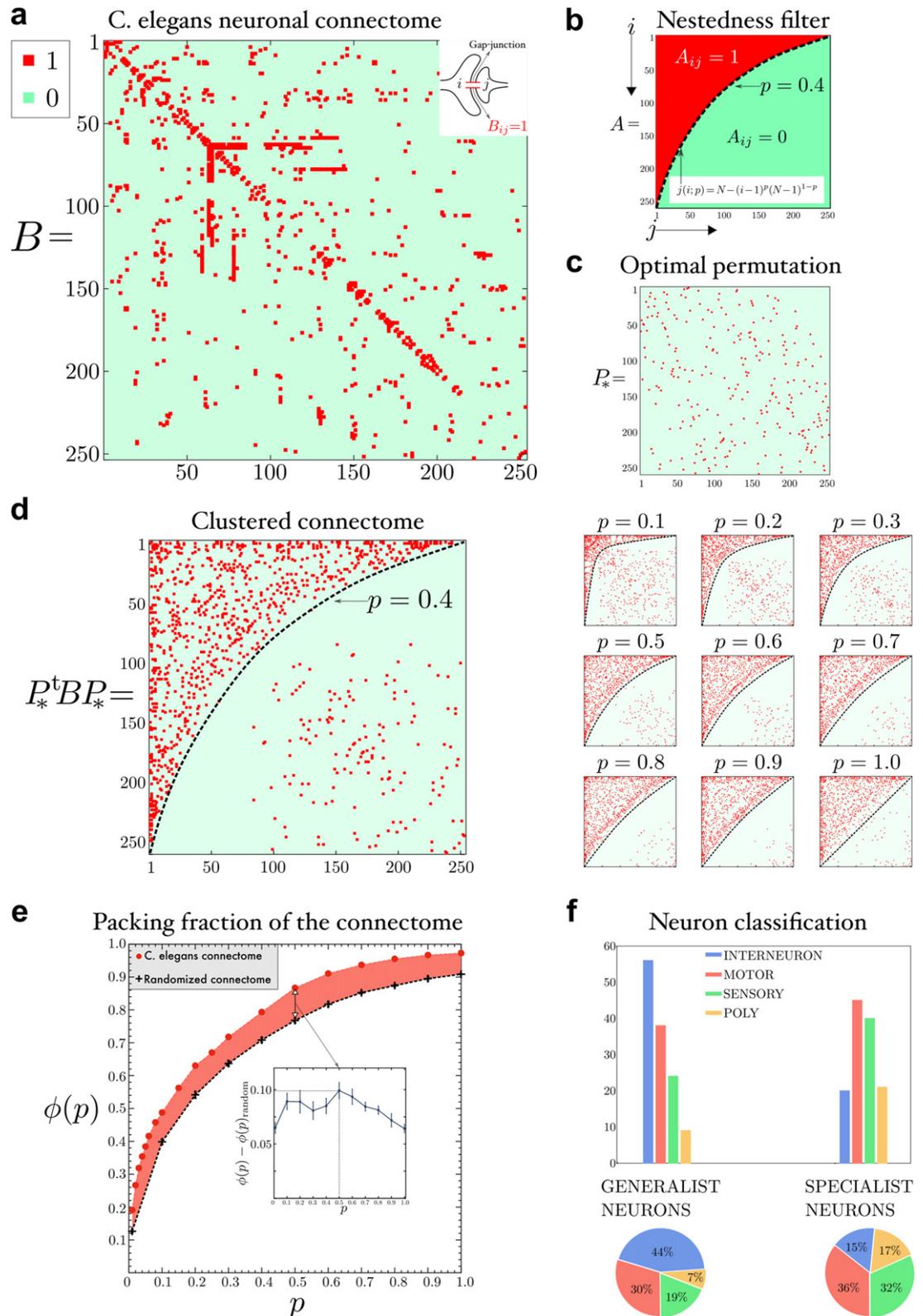


Figure 2. Clustering the *C. elegans* connectome through optimal permutations. (a) The adjacency matrix B of the *C. elegans* gap-junction connectome from reference [16]. The matrix is binary so its entries take two possible values: $B_{ij} = 1$ if a gap-junction exists between neurons i and j , and $B_{ij} = 0$ if not. (b) The nestedness filter A used to cluster the adjacency matrix B . Matrix A is a binary matrix having entries $A_{ij} = 1$ for $j \leq f(i; p) = N - (i - 1)^p (N - 1)^{1-p}$ (corresponding to the red area extending from the upper left corner to the black dashed line defined by the equation $j = f(i; p)$); and $A_{ij} = 0$ for $j > f(i; p)$ (corresponding to the complementary light-green area). We choose the nestedness exponent $p = 0.4$. (c) The optimal permutation matrix P_* obtained by solving the OPP with the matrices B and A shown in (a) and (b) respectively. (d) Left side: the clustered adjacency matrix obtained from B by applying a similarity transformation with the optimal permutation matrix P_* found in (c), that is $P_*^t B P_*$. Right side: clustered adjacency matrices obtained with nine different filters having nestedness exponents $0.1 \leq p \leq 1.0$. (e) The packing fraction of *C. elegans* connectome $\phi(p)$ (red dots), defined by equation (26), as a function of the nestedness exponent p , as compared to the average packing fraction $\phi_{\text{ran}}(p)$ (black crosses) of a randomized connectome with the same degree sequence (error bars are s.e.m. over 10 realizations of the configurational model). The inset shows the difference $\phi(p) - \phi_{\text{ran}}(p)$ as a function of p , that has a maximum equal to ~ 0.1 for $p = 0.5$. (f) Classification of neurons into generalists and specialists as explained in the main text.

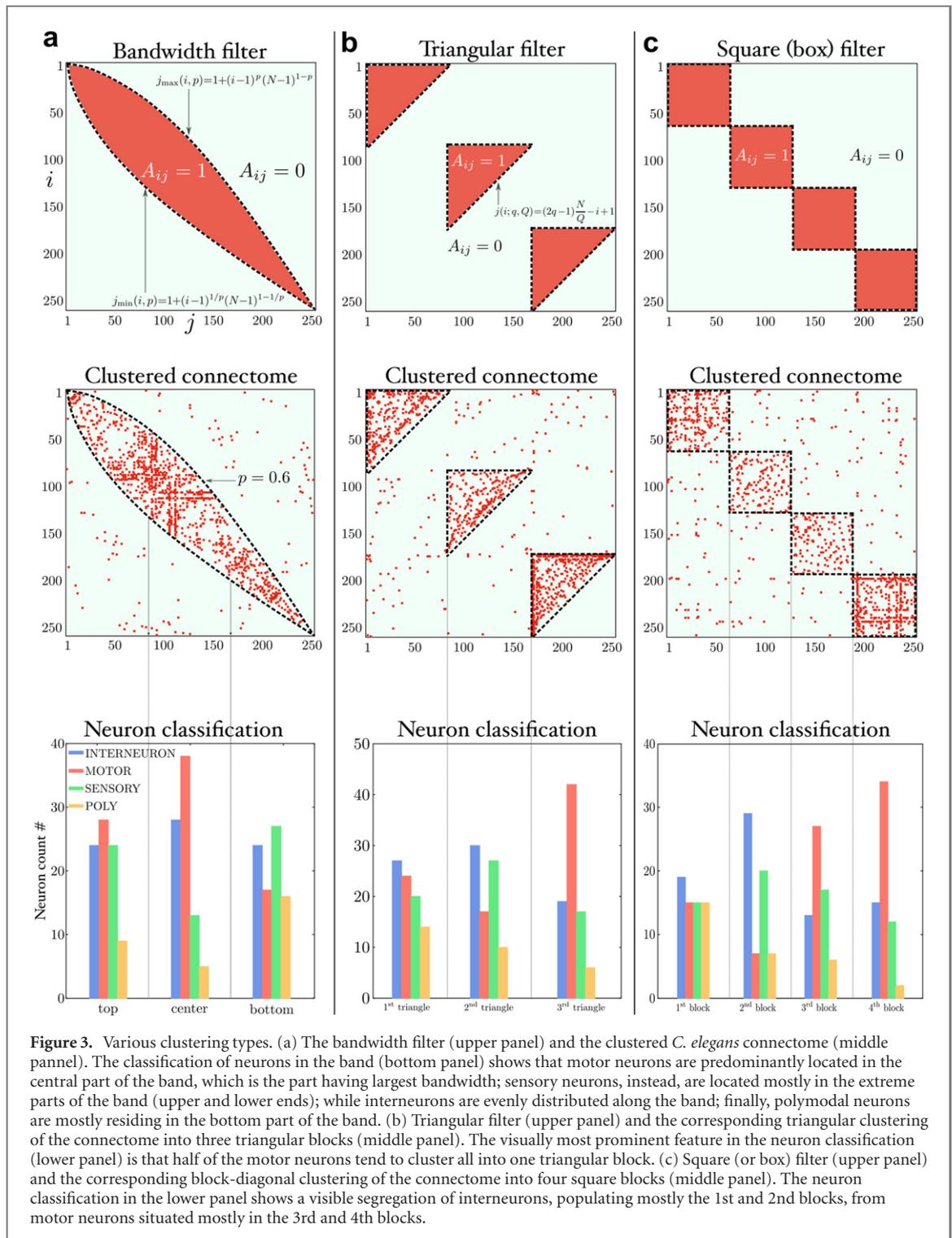
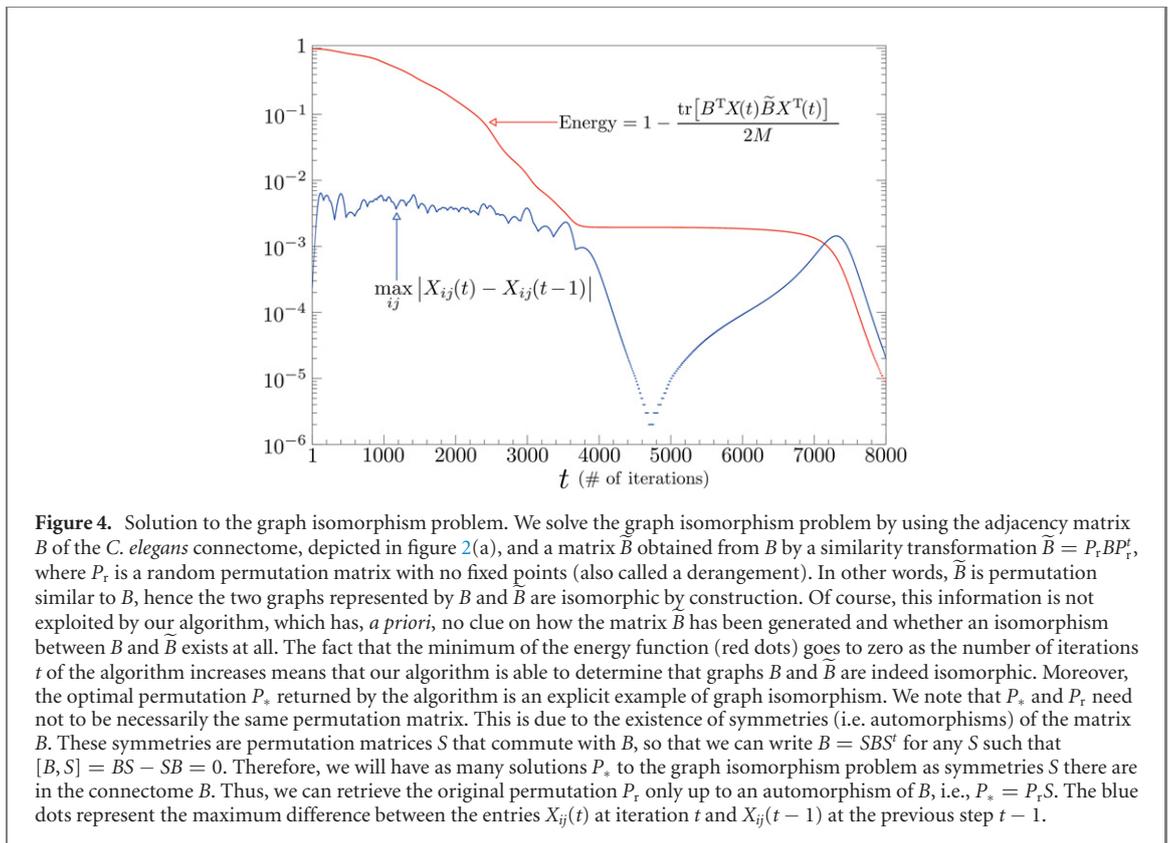


Figure 3. Various clustering types. (a) The bandwidth filter (upper panel) and the clustered *C. elegans* connectome (middle panel). The classification of neurons in the band (bottom panel) shows that motor neurons are predominantly located in the central part of the band, which is the part having largest bandwidth; sensory neurons, instead, are located mostly in the extreme parts of the band (upper and lower ends); while interneurons are evenly distributed along the band; finally, polymodal neurons are mostly residing in the bottom part of the band. (b) Triangular filter (upper panel) and the corresponding triangular clustering of the connectome into three triangular blocks (middle panel). The visually most prominent feature in the neuron classification (lower panel) is that half of the motor neurons tend to cluster all into one triangular block. (c) Square (or box) filter (upper panel) and the corresponding block-diagonal clustering of the connectome into four square blocks (middle panel). The neuron classification in the lower panel shows a visible segregation of interneurons, populating mostly the 1st and 2nd blocks, from motor neurons situated mostly in the 3rd and 4th blocks.

(or box) filter [6] in figure 3(c), whose mathematical properties are discussed in appendix C and in the caption of figure 3.

To conclude this section, we notice that if A represents itself the graph of a network, then the OPP is equivalent to the graph isomorphism problem [25], as exemplified in figure 4. Furthermore, the graph isomorphism becomes the graph automorphism problem in the special case $A = B$. The graph automorphism problem, in turn, is equivalent to the problem of minimizing the norm of the commutator $E(P) = \|[A, P]\|^2$ (with P different from the identity matrix). The solution P_* is called a ‘symmetry of the network’ if $E(P_*) = 0$, or a ‘pseudosymmetry’ if the condition $E(P_*) > 0$ holds true [26].



4. Conclusion

Our analytic and algorithmic results for clustering a matrix by solving the OPP reveal their importance in that their essential features are not contingent on a special form of the input matrix, nor on special assumptions concerning the filter matrices.

The saddle point solution presented here can be thought of as the first order term of a perturbative expansion of the exact solution. Therefore, our framework allows one also to systematically improve over the saddle point solution by computing more terms of the perturbative expansion, that are analogous to loop corrections to the tree approximation in statistical field theory [10].

Finally, we observe that our framework can be easily extended and the corresponding algorithm readily applied to cluster rectangular matrices. For example, to cluster a $N \times M$ matrix B_{ia} according to a $N \times M$ filter A_{ia} , we define an energy function $E(P, Q) = -\text{Tr}(B^t P A Q^t)$, which is a function of two permutations matrices: a $N \times N$ permutation matrix P and a $M \times M$ permutation matrix Q , and then minimize it simultaneously over P and Q along the same lines explained in this work.

Acknowledgments

We thank Dr Manuel Zimmer for providing the time series to calculate the correlation matrix shown in figure 1(a). This work was supported by AFOSR: Grant FA9550-21-1-0236.

Data availability statement

Data that support the findings of this study are publicly available at the Wormatlas database at <https://wormatlas.org>. The source code of the clustering algorithm can be downloaded from GitHub at <https://github.com/flavianoM/Clustering-Matrices-through-optimal-permutations>.

Author contributions

FM designed research, developed the theory, analyzed the data, and wrote the manuscript.

Conflict of interest

The author declares no competing interests.

Appendix A. Projected free-energy

In this appendix, we derive the expression of the free-energy projected onto the global constraints (i.e. the free-energy evaluated on the solution of the saddle point equations) as given in equation (23). We start from the unconstrained free-energy given by equation (13), that we rewrite below for simplicity

$$F(X, Y, z) = E(X) + \frac{1}{\beta} \text{Tr}(XY^t) - \frac{1}{\beta} \sum_i \log \sum_j e^{Y_{ij}} + \frac{1}{\beta} \sum_j z_j \left(\sum_i X_{ij} - 1 \right). \quad (\text{A1})$$

At the saddle point, X satisfies the constraints

$$\sum_i X_{ij} = 1, \quad (\text{A2})$$

and therefore equation (A1) becomes

$$F(X, Y) = E(X) + \frac{1}{\beta} \text{Tr}(XY^T) - \frac{1}{\beta} \sum_i \log \sum_j e^{Y_{ij}}. \quad (\text{A3})$$

Let us now consider the term $\sum_i \log \sum_j e^{Y_{ij}}$, that we can rewrite as

$$\sum_i \log \sum_j e^{Y_{ij}} = \sum_i \left(\sum_j X_{ij} \right) \log \sum_j e^{Y_{ij}}, \quad (\text{A4})$$

where we have used the fact that, at the saddle point, X satisfies $\sum_j X_{ij} = 1$. Next, we use the following saddle point equation relating X and Y :

$$\frac{e^{Y_{ij}}}{\sum_k e^{Y_{ik}}} = X_{ij} \quad (\text{A5})$$

to write

$$\sum_k e^{Y_{ik}} = \frac{e^{Y_{ij}}}{X_{ij}}. \quad (\text{A6})$$

Inserting equation (A6) into equation (A4) we obtain

$$\sum_{ij} X_{ij} \log \sum_j e^{Y_{ij}} = \sum_{ij} X_{ij} Y_{ij} - \sum_{ij} X_{ij} \log X_{ij} = \text{Tr}(XY^t) - \sum_{ij} X_{ij} \log X_{ij}. \quad (\text{A7})$$

Inserting equation (A7) into equation (A3) we obtain

$$F(X) = E(X) + \frac{1}{\beta} \sum_{ij} X_{ij} \log X_{ij}, \quad (\text{A8})$$

which is the free-energy given in equation (23) of the main text.

Appendix B. Algorithm to solve the saddle point equations and find the optimal permutation

In this appendix, we explain the algorithm to find the doubly-stochastic matrix X which solves the saddle point equation (15). First, we must ensure that our algorithm finds a solution that lies on a vertex of the hypercube $[0, 1]^{N^2}$. To address this issue let us consider the Hessian $H_{ij,kl}$ given by equation (24). If the Hessian has a positive eigenvalue in the interior of the hypercube, than there can be a minimum there. Therefore, to prevent minima from lying inside the hypercube we must ensure that all eigenvalues of the Hessian are negative inside the hypercube. To achieve this, we need to add an additional term to the free-energy as follows

$$F(X) \rightarrow F_\epsilon(X) = F(X) - \frac{1}{2} |\epsilon| \text{Tr}(XX^t), \quad (\text{B1})$$

so that the Hessian becomes

$$H_{ij;kl} \rightarrow H_{ij;kl} - |\epsilon|. \tag{B2}$$

Thus, by taking $|\epsilon|$ sufficiently large we can ensure that $F_\epsilon(X)$ will not have a minimum inside the hypercube.

The regularization of the free-energy given in equation (B1) can be understood also from a different perspective. Specifically, let us consider the energy function of the original problem $E(P) = \|PA - BP\|^2$ and expand the Frobenius norm as

$$E(P) = \|PA - BP\|^2 = \text{Tr}(PAA^tP^t) + \text{Tr}(BPP^tB^t) - 2 \text{Tr}(PAP^tB^t). \tag{B3}$$

Since a permutation matrix is an orthogonal matrix, i.e. $P^t = P^{-1}$, the first two terms on the right-hand side of equation (B3) do not depend on P ; hence the problem of minimizing $E(P)$ is equivalent to the problem of minimizing the term $-2 \text{Tr}(PAP^tB^t)$. The crux of the matter is that, by approximating P with a doubly-stochastic matrix X , we lose orthogonality, meaning that $X^tX \neq I$. Therefore, the $|\epsilon|$ term in equation (B1) can be also interpreted as a Lagrange multiplier enforcing the orthogonality constraint $\text{Tr}(XX^t) = N$ (the term $|\epsilon|N/2$ is omitted from $F_\epsilon(X)$ because it is a constant independent of X).

Initialization of the algorithm

- We initialize $X_{ij}^{(t=0)} \equiv X_{ij}^{(0)}$ at the beginning of the algorithm to a uniform random matrix as: $X_{ij}^{(0)} = r$, where r is uniformly distributed in $(0, 2/N]$.
- We initialize the value of $|\epsilon|$ to be

$$|\epsilon| = \left(\max_i |\lambda_i^{(A)}| \right) \left(\max_i |\lambda_i^{(B)}| \right), \tag{B4}$$

at the beginning of the algorithm, and then decrease it as $\epsilon \rightarrow \epsilon - 1$ until $\epsilon = 0$, after each execution of the following routine.

Steepest descent

We split the saddle-point equations into two pieces as follows

$$\begin{aligned} h_{ij} &= \frac{\beta}{2}(BXA^t + B^tXA) + \beta|\epsilon|X_{ij}, \\ X_{ij} &= U_i e^{h_{ij}} V_j. \end{aligned} \tag{B5}$$

We can think of the equation involving h_{ij} as the stationary state of the following dynamical system

$$\frac{dh_{ij}}{dt} = -\eta \left\{ h_{ij}(t) - \left[\frac{\beta}{2}(BX(t)A^t + B^tX(t)A) + \beta|\epsilon|X_{ij}(t) \right] \right\}, \tag{B6}$$

which in discrete time becomes

$$h_{ij}(t + 1) = (1 - \eta)h_{ij}(t) + \eta \left[\frac{\beta}{2}(BX(t)A^t + B^tX(t)A) + \beta|\epsilon|X_{ij}(t) \right]. \tag{B7}$$

Equation (B7) represents the updating rule for the auxiliary variables $h_{ij}(t)$, and depends on a small parameter $\eta > 0$. In our experiments we set $\eta = 10^{-3}$.

Having updated $h_{ij}(t + 1)$, we move to update X_{ij} by solving the following coupled equations for U_i and V_j

$$\begin{aligned} U_i &= \left(\sum_j e^{h_{ij}(t+1)} V_j \right)^{-1}, \\ V_j &= \left(\sum_i U_i e^{h_{ij}(t+1)} \right)^{-1}, \end{aligned} \tag{B8}$$

and then setting

$$X_{ij}(t + 1) = U_i e^{h_{ij}(t+1)} V_j. \tag{B9}$$

To monitor the convergence of the algorithm, we compute at each step the maximum change in the entries X_{ij} , i.e.

$$\text{diff} \equiv \max_{ij} |X_{ij}(t + 1) - X_{ij}(t)|, \tag{B10}$$

and we return X when the following criterion is met

$$\text{IF diff} < \delta \rightarrow \text{RETURN } X_{ij}. \tag{B11}$$

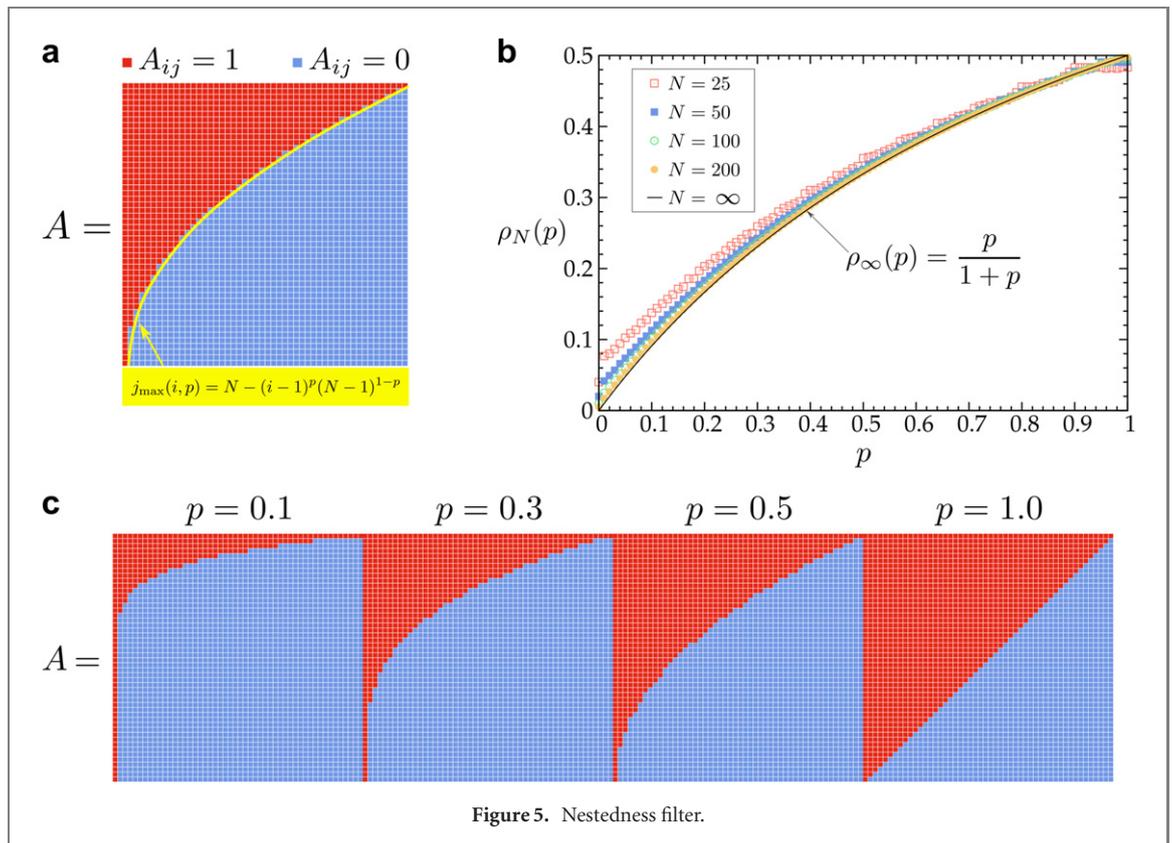


Figure 5. Nestedness filter.

In our experiments we choose $\delta = 10^{-5}$.

Moreover, in all our numerical experiments, we observed that for sufficiently low temperature (in practice for $\beta \geq 10$) the output matrix X_{ij} is a permutation matrix, i.e. it corresponds to a vertex of the hypercube $[0, 1]^{N^2}$.

Appendix C. Filter matrices

In this appendix, we describe the filter matrices A that we used in our clustering experiments discussed in the main text.

C1. Nestedness filter

The nestedness filter is described by a matrix A whose nonzero entries A_{ij} are equal to 1 when the following condition is satisfied:

$$A_{ij} = 1 \quad \text{for} \quad 1 \leq i \leq N \quad \text{and} \quad j \in [1, j_{\max}(i, p)], \quad (C1)$$

where $j_{\max}(i, p)$ is given by:

$$j_{\max}(i, p) = N - (i - 1)^p(N - 1)^{1-p}, \quad (C2)$$

as shown in figure 5(a). The parameter $p \in [0, 1]$ quantifies the nestedness of the matrix A . Specifically, low values of p correspond to a matrix A with a highly nested structure. Opposite to this, large values of p , i.e. $p \sim 1$, describe profiles of low nestedness, as depicted in figures 5(a) and (c).

The density $\rho(p)$ of the filter matrix A defined by equation (C1) is given by

$$\rho(p) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^{j_{\max}(i,p)} 1, \quad (C3)$$

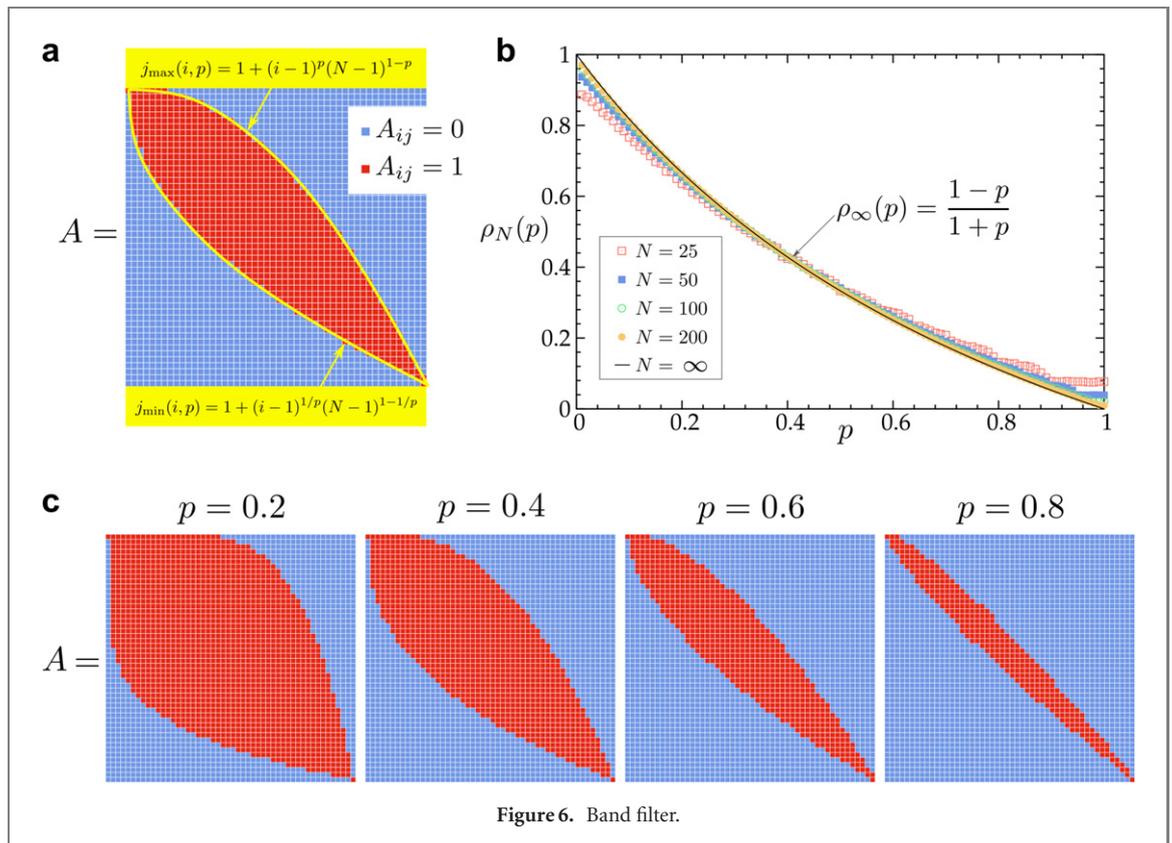


Figure 6. Band filter.

which, in the limit $N \rightarrow \infty$, becomes:

$$\rho(p) = \frac{p}{1+p}. \tag{C4}$$

The finite N behavior of $\rho(p)$ together with its $N \rightarrow \infty$ limit is shown in figure 5(b).

C2. Band filter

The band filter is a matrix A whose entries $A_{ij} \in \{0, 1\}$ are defined by

$$A_{ij} = \begin{cases} 1 & \text{for } 1 \leq i \leq N \\ & j_{\min}(i,p) \leq j \leq j_{\max}(i,p) \\ 0 & \text{otherwise} \end{cases}, \tag{C5}$$

where

$$\begin{aligned} j_{\min}(i,p) &= 1 + (i-1)^{1/p}(N-1)^{1-1/p}, \\ j_{\max}(i,p) &= 1 + (i-1)^p(N-1)^{1-p}, \end{aligned} \tag{C6}$$

where p is a parameter that controls the width of the band, hence we call p the **bandwidth** exponent. The band filter in equation (C5) has nonzero entries comprised in a band delimited by $j_{\min}(i,p)$ and $j_{\max}(i,p)$ for $i = 1, \dots, N$. The density $\rho(p)$ of A is defined as the fraction of entries contained inside the band:

$$\rho(p) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=j_{\min}(i,p)}^{j_{\max}(i,p)} 1. \tag{C7}$$

For $N \rightarrow \infty$, the density $\rho(p)$ evaluates

$$\rho(p) = \frac{1-p}{1+p}. \tag{C8}$$

The finite N behavior of $\rho(p)$ along with the $N \rightarrow \infty$ limit given by equation (C8) are shown in figure 6(b).

A useful quantity to characterize the shape of the band filter is the **bandwidth** $b(p)$, which is defined by

$$b(p) = \max_i [j_{\max}(i,p) - j_{\min}(i,p)]. \tag{C9}$$

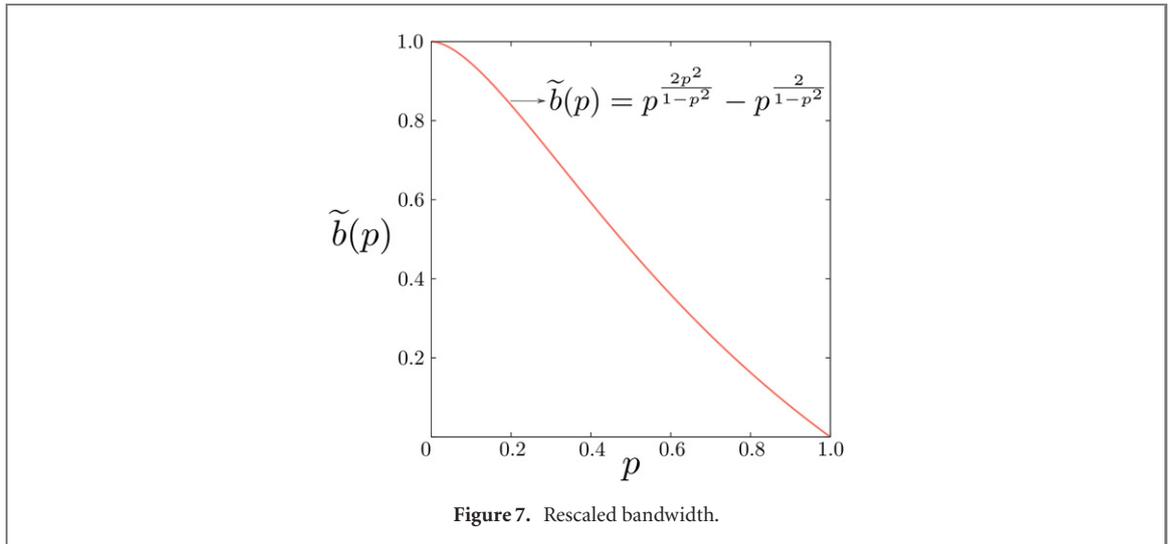


Figure 7. Rescaled bandwidth.

Let us define the rescaled coordinate x taking values in the range $x \in [0, 1]$ as:

$$x = \frac{i - 1}{N - 1}, \tag{C10}$$

whereby we can write the difference $j_{\max} - j_{\min}$ as

$$j_{\max}(i, p) - j_{\min}(i, p) = (N - 1)(x^p - x^{1/p}). \tag{C11}$$

Next we define the rescaled bandwidth $\tilde{b}(p)$ as

$$\tilde{b}(p) = \frac{b(p)}{N - 1}. \tag{C12}$$

Thus, in the large N limit we can approximate x to a continuous variable and thus estimate $\tilde{b}(p)$ as follows:

$$\tilde{b}(p) = x_*(p)^p - x_*(p)^{1/p}, \tag{C13}$$

where $x_*(p)$ is the solution to the following equation:

$$\left. \frac{d}{dx}(x^p - x^{1/p}) \right|_{x=x_*(p)} = 0, \tag{C14}$$

that is:

$$x_*(p) = p^{\frac{2p}{1-p^2}}. \tag{C15}$$

Substituting equation (C15) into equation (C13) we obtain the explicit form of the rescaled bandwidth as a function of p

$$\tilde{b}(p) = p^{\frac{2p^2}{1-p^2}} - p^{\frac{2}{1-p^2}}, \tag{C16}$$

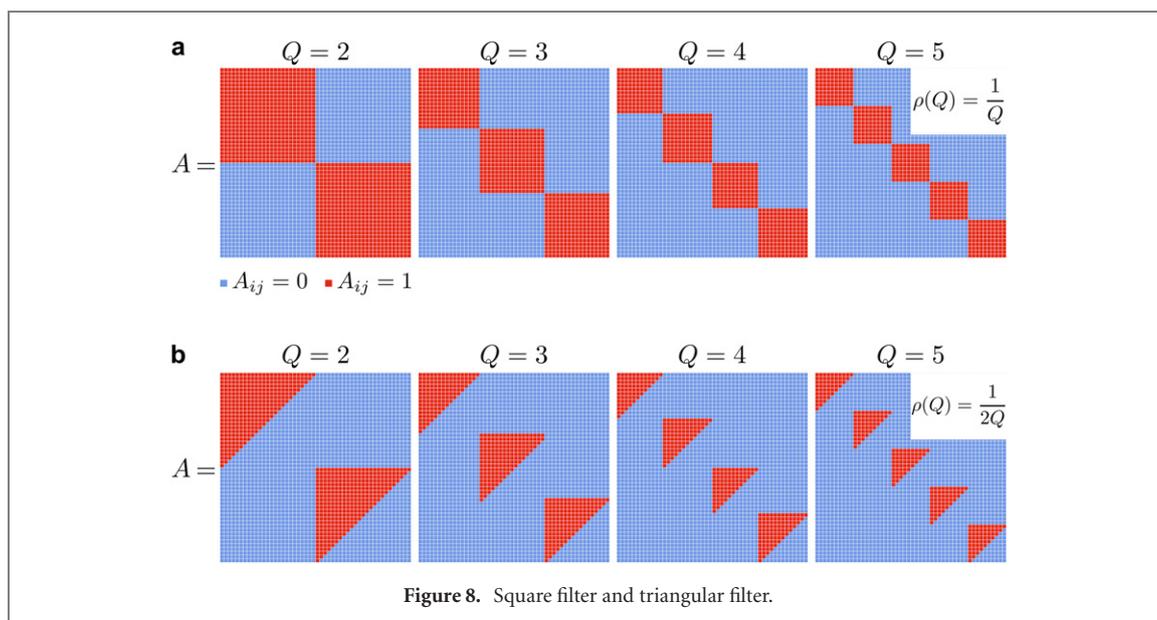
which is shown in figure 7.

C3. Square filter and triangle filter

The square filter A is shown in figure 8(a) and is parameterized by a number Q representing the number of blocks the matrix A is divided into. The size of each block is $\frac{N}{Q} \times \frac{N}{Q}$. The Q square blocks are arranged along the main diagonal. Mathematically, the entries A_{ij} of A are defined to be 0 or 1 by

$$A_{ij} = 1 \quad \text{if} \quad \begin{cases} i \in \left[1 + (q - 1)\frac{N}{Q}, q\frac{N}{Q} \right] \\ \text{and} \\ j \in \left[1 + (q - 1)\frac{N}{Q}, q\frac{N}{Q} \right] \end{cases}, \quad q = 1, \dots, Q \tag{C17}$$

$$A_{ij} = 0 \quad \text{otherwise}$$



The density $\rho(Q)$ is easy to calculate and evaluates:

$$\rho(Q) = \frac{1}{Q}. \tag{C18}$$

The triangle filter is shown in figure 8(b) and is parameterized by a number Q representing the number of equal sized triangular blocks arranged along the main diagonal of the matrix. Mathematically, the entries A_{ij} of A are defined to be 0 or 1 by

$$A_{ij} = 1 \quad \text{if} \quad \begin{cases} i \in \left[1 + (q-1)\frac{N}{Q}, q\frac{N}{Q} \right] \\ \text{and} \\ j \in \left[1 + (q-1)\frac{N}{Q}, (2q-1)\frac{N}{Q} - i + 1 \right] \end{cases}, \quad q = 1, \dots, Q \tag{C19}$$

$$A_{ij} = 0 \quad \text{otherwise}$$

The density $\rho(Q)$ of the triangle filter equals to

$$\rho(Q) = \frac{1}{2Q}. \tag{C20}$$

ORCID iDs

Flaviano Morone  <https://orcid.org/0000-0002-8830-1179>

References

- [1] Golub G H and Van Loan C F 1996 *Matrix Computations* 3rd edn (Baltimore, MD: John Hopkins University Press)
- [2] D’haeseleer P 2005 How does gene expression clustering work? *Nat. Biotechnol.* **23** 1499–501
- [3] Kato S, Kaplan H S, Schrödel T, Skora S, Lindsay T H, Yemini E, Lockery S and Zimmer M 2015 Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans* *Cell* **163** 656–69
- [4] Newman M 2018 *Networks* (Oxford: Oxford University Press)
- [5] ISO/IEC Standard 18004 2006 *Information Technology—Automatic Identification and Data Capture Techniques—QR Code 2005 Bar Code Symbol Specification* (International Organisation for Standardization)
- [6] Jain A K and Dubes R C 1988 *Algorithms for Clustering Data* (Englewood Cliffs, NJ: Prentice-Hall)
- [7] Koopmans T C and Beckmann M 1957 Assignment problems and the location of economic activities *Econometrica* **25** 53–76
- [8] Papadimitriou C H and Steiglitz K 1998 *Combinatorial Optimization: Algorithms and Complexity* (North Chelmsford: Courier)
- [9] WormAtlas Altun Z F, Herndon L A, Wolkow C A, Crocker C, Lints R and Hall D H 2002–2019 <https://wormatlas.org>
- [10] Zinn-Justin J 2002 *Quantum Field Theory and Critical Phenomena* 4th edn (Oxford: Oxford University Press)
- [11] Sinkhorn R 1964 A relationship between arbitrary positive matrices and doubly stochastic matrices *Ann. Math. Stat.* **35** 876–9
- [12] Sinkhorn R and Knopp P 1967 Concerning nonnegative matrices and doubly stochastic matrices *Pac. J. Math.* **21** 343–8
- [13] Tacchella A, Cristelli M, Caldarelli G, Gabrielli A and Pietronero L 2012 A new metrics for countries’ fitness and products’ complexity *Sci. Rep.* **2** 723

- [14] Kirkpatrick S, Gelatt C D Jr and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671–80
- [15] Kosowsky J J and Yuille A L 1994 The invisible hand algorithm: solving the assignment problem with statistical physics *Neural Netw.* **7** 477–90
- [16] Varshney L R, Chen B L, Paniagua E, Hall D H and Chklovskii D B 2011 Structural properties of the *Caenorhabditis elegans* neuronal network *PLoS Comput. Biol.* **7** e1001066
- [17] Patterson B D and Atmar W 1986 Nested subsets and the structure of insular mammalian faunas and archipelagos *Biol. J. Linn. Soc.* **28** 65–82
- [18] Bascompte J, Jordano P, Melián C J and Olesen J M 2003 The nested assembly of plant-animal mutualistic networks *Proc. Natl. Acad. Sci. USA* **100** 9383–7
- [19] Staniczenko P P A, Kopp J C and Allesina S 2013 The ghost of nestedness in ecological networks *Nat. Commun.* **4** 1931
- [20] König M D, Tessone C J and Zenou Y 2014 Nestedness in networks: a theoretical model and some applications *Theor. Econ.* **9** 695–752
- [21] Mariani M S, Ren Z-M, Bascompte J and Tessone C J 2019 Nestedness in complex networks: observation, emergence, and implications *Phys. Rep.* **813** 1–90
- [22] Kaplan H S, Salazar Thula O, Khoss N and Zimmer M 2019 Nested neuronal dynamics orchestrate a behavioral hierarchy across timescales *Neuron* **105** 562–76
- [23] Morone F, Del Ferraro G and Makse H A 2019 The k-core as a predictor of structural collapse in mutualistic ecosystems *Nat. Phys.* **15** 95–102
- [24] Chinn P Z, Chvátalová J, Dewdney A K and Gibbs N E 1982 The bandwidth problem for graphs and matrices—a survey *J. Graph Theory* **6** 223–54
- [25] Babai L 2016 Graph isomorphism in quasipolynomial time *STOC '16: Proc. 48th Annual ACM Symp. on Theory of Computing* pp 684–97
- [26] Morone F and Makse H 2019 Symmetry group factorization reveals the structure-function relation in the neural connectome of *Caenorhabditis elegans* *Nat. Commun.* **10** 4961