

Reconstructing Cortical Networks: Case of Directed Graphs with High Level of Reciprocity

Tamás Nepusz, László Négyessy, Gábor Tusnány, and Fülöp Bazsó

Abstract The problem of prediction of yet uncharted connections in the large scale network of the cerebral cortex is addressed. Our approach was determined by the fact that the cortical network is highly reciprocal although directed, i.e. the input and output connection patterns of vertices are slightly different. In order to solve the problem of predicting missing connections in the cerebral cortex, we propose a probabilistic method, where vertices are grouped into two clusters based on their outgoing and incoming edges, and the probability of a connection is determined by the cluster affiliations of the vertices involved. Our approach allows accounting for differences in the incoming and outgoing connections, and is free from assumptions about graph properties. The method is general and applicable to any network for which the connectional structure is mapped to a sufficient extent. Our method allows the reconstruction of the original visual cortical network with high accuracy, which was confirmed after comparisons with previous results. For the first time, the effect of extension of the visual cortex was also examined on graph reconstruction after complementing it with the subnetwork of the sensori-

Tamás Nepusz

KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences, H-1525 Budapest, P.O. Box 49., Hungary and Budapest University of Technology and Economics, Department of Measurement and Information Systems, H-1521 Budapest, P.O. Box 91., Hungary, e-mail: ntamas@rmki.kfki.hu

László Négyessy

Neurobionics Research Group, Hungarian Academy of Sciences – Péter Pázmány Catholic University – Semmelweis University, H-1094 Budapest, Tűzoltó u. 58., Hungary

Gábor Tusnány

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Reáltanoda u. 13-15., 1053 Budapest, Hungary

Bazsó Fülöp

KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences, H-1525 Budapest, P.O. Box 49., Hungary and Polytechnical Engineering College Subotica, Marka Oreškovića 16, 24000 Subotica, Serbia, e-mail: bazso@mail.kfki.hu

motor cortex. This additional connectional information further improved the graph reconstruction. One of our major findings is that knowledge of definitely nonexistent connections may significantly improve the quality of predictions regarding previously uncharted edges as well as the understanding of the large scale cortical organization.

1 Introduction

The cerebral cortex is probably the most prominent example of a natural information processing network. It is therefore of major importance to learn how this network is organized. At the lowest level, the cortical network is composed of physically (i.e. via chemical and electrical synapses) connected nerve cells. (The chemical synapse is the dominant type and it is rectifying in contrast to the electrical synapse, which allows bi-directional interactions between the neurons). The cortex, in general (ignoring species and areal density differences [16]) consists of approximately 10^{10} nerve cells, each receiving numerous connections of order 10^3 (up to about 10^4) [7]. However, these data do not necessarily imply a homogenous degree distribution as the diverse types of neurons could form specific sub-networks. Based on functional constraints and axonal wiring economy, Buzsáki et al. [10] proposed a scale-free-like, small world architecture for the network of inhibitory local circuit neurons consisting approximately 20% of the whole neuron population. In fact, the significance of such a diversity of neurons is presently unknown, especially in the case of the pyramidal cells, which is the principal cell type of the cerebral cortex (making up the remaining 80%) [16]. Pyramidal cells, considered excitatory in nature, form long distance connections both within the cortex and with subcortical structures. However, most of the synaptic contacts are formed locally within a short distance, and it is unclear how the cortical network is organized at the neuron level [18]. Considering anatomical and physiological data, Tononi et al. [66] outlined a network architecture, which suitably performs segregation and integration, the fundamental functions of the central nervous system. Integration is achieved by connections between clusters of neurons, representing functionally specialized units, which are formed by dense local connectivity [66]. Using mutual information as a measure of integration, it was shown that the proposed network exhibited high complexity, significantly differing from random and regular lattice networks characterized by low measure of complexity [66]. Differences between cortical and random networks were also pointed out by Négyessy et al. [46], although on the level of cortical areas instead of single neurons. On the other hand, based on estimates of the spreading function, Bienenstock [6] showed that the graph of cortical neurons has a high dimensionality close to that of an Erdős-Rényi random graph of similar size. This assumption is consonant with Szentágothai's notion of quasi-randomness in neuronal connectivity [64].

The neuron doctrine (stating that nerve cells are the developmental, structural and functional units of the central nervous system) has been challenged by arguing that populations of neurons function as units [5, 8, 19, 23, 44, 63, 66]. This perspective is in close agreement with the so-called columnar or modular organization of cortical structure and function as proposed by Mountcastle [44] and Szentágothai [63]. Accordingly, the cortex is usually viewed as a two dimensional sheet composed of functional modules with a diameter of 250–500 μm , arranged perpendicularly to the surface and spanning the layers to the depth of the cortex. Although it is hard to estimate due to the different types (and size) of columns [11, 44], a network of such modules would form a graph with millions of vertices in case of humans. Unfortunately, such a network would be hard to draw because apart from some cortical regions and specific columns (e.g., [4, 9, 35, 56, 57, 61]) the interconnections among these modules or cell clusters are obscure. In addition, functional modules may not be fixed structures; they could dynamically change their extension via neuronal plasticity (e.g., [11, 31]). It is noteworthy that the minicolumnar organization apparently can not resolve the problem of defining structural and functional cortical units, as momentarily the minicolumn seems to be a similarly vague concept as the column [18, 55].

At a higher organizational level, the cortex is composed of a set (about a hundred, roughly four orders of magnitude less than the number of columnar modules in the human) of structurally and functionally specialized regions or areas with highly variable shapes and sizes [67]. This level of organization is of great interest because the available anatomical and imaging (fMRI, PET, EEG, MEG) techniques made it possible to investigate the network of cortical areas (hence neuro-cognitive functions) [1, 8, 29, 41, 50]. Most of our knowledge about this large-scale cortical network comes from studies charting the neuronal connections between cortical areas. Since the use of sensitive and powerful tract tracing techniques is not feasible in humans, the neural connections among the areas (“anatomical connectivity”, [59]) have been studied intensely in non-human primates, especially in the macaque, which serves as a model of the human cortex [67]. Large collections of such data are available at the CoCoMac database [37] and for updates, one may search PubMed [52]. Although the areas are connected to each other via varying density of bundles of axonal processes [27] in a complicated laminar and topographical pattern [20, 54], the network of areas is usually represented in binary form considering only the knowledge of the existence of a connection between the areas [58]. Even such a simplification allowed the description of the fundamental properties of the network of cortical areas, e.g., its small-world like characteristics and hierarchical organization [28, 32, 59]. However, Kötter and Stephan [38] have pointed out that the lack of information about connectivity can hinder the understanding of important features of the cortical network.

The cortical network is directed, as long-range connections between the areas end up in chemical synapses, but strongly reciprocal (reciprocity reaches 80%) [20]. From the graph theoretic point of view, the high level of reci-

procuity presents an obstacle by obscuring directedness in the network. The global edge density is roughly 0.2–0.4 [58]. A granular organization representing functional segregation and integration is prevalent also at this large scale [26, 30, 45, 47, 68], resulting in the density of connections going up to 0.6 or more within the clusters [21]. The other major characteristic that makes the cortex a small-world network is that its clustered organization is accompanied by a short average path length, roughly between 2 and 3 [32, 59]. Because it is reasonable to assume that a considerable part of the large-scale cortical network is still unknown, the identification of the key topological features that characterize this network, i.e. understanding its organizational principles, remained an open issue [21, 30, 33, 34, 38, 51, 60]. A practical way of approaching this problem is to check how exactly the network can be reconstructed by using a given index or network measure [21, 30]. This approach also has the interesting consequence of predicting missing data, which can be verified experimentally. The two studies published up to now present data on such predictions of yet unknown connections in the cortex [21, 30]. The results of these studies (especially those by Costa et al. [21], who investigated a broad set of measures) suggest that connectional similarity of the areas is a good predictor in reconstructing the original cortical network. However, they also report a relatively large number of violations, where known existent connections were predicted as nonexistent in the reconstructed graphs and vice versa [21, 30]. This suggests that using other approaches could result in better reconstruction of the cortical network. The aim of the present study was therefore to find a reconstruction algorithm that predicts the large-scale cortical network more accurately, i.e. with fewer violations. By considering the similarity of the connections of the individual areas, our approach is reminiscent to that used by the previous analyses [21, 30]. However, there are substantial differences as well, especially the fact that we use a stochastic method which is able to take into account the amount of uncertainty present in the data being analyzed. Furthermore, contrary to the previous studies [21, 30], where the in- and outputs are either taken into account separately [30] or the network was symmetrized prior to analysis [21], our approach is principally dependent on the combination of the areas’ in- and output pattern. Notably, considering the similarity of the in- and output pattern as the result of the high number of reciprocated links, a stochastic approach seems advantageous. Finally, in contrast to Jouve et al. [30], who assumed that a large number of indirect connections of path length 2 is suggestive of the existence of a direct link between the areas, our method is free of such assumptions.

2 Methods

In this section, we introduce a simple stochastic graph model based on vertex types and connection probabilities depending on them. From now on, we call

this the *preference model*. We will discuss a greedy and a Markov chain Monte Carlo (MCMC) method for fitting the parameters of the preference model to a network being studied. The MCMC method shows the rapid mixing property. In Section 3, we will employ these methods to reconstruct the cortical network and to predict previously unknown connections.

2.1 *General remarks on the method*

The problem we would like to solve and the proposed method are not cortex specific, though the data on which we operate is. As data collection and mapping are necessarily partial due to unavoidable observational errors, our method offers the possibility to map interactions, connections, influences based on the previous knowledge. Given the rough, but in principle correct summary of such information in the form of appropriate graph model, one may refine the knowledge of underlying graph representation to some extent. Applications and extensions of the solution we propose are straightforward to apply to any other network, with appropriate caution. The main assumptions underlying our approach are as follows: the number of nodes is known in advance, we only wish to predict previously uncharted edges, the majority of the edges are known, yet a large number of undetected edges are possible, at least in principle. As with most problems involving prediction, it is relatively simple to create a model performing slightly better than random tossing, but increasing prediction accuracy is a difficult problem. Our approach to edge prediction is inspired by one of the most influential results of graph theory by Endre Szemerédi [36, 62, 65], which became known as Szemerédi’s regularity lemma. Loosely speaking, the regularity lemma states that the structures of very large graphs can be described by assigning vertices to a small number of almost equally sized groups and specifying the connection probabilities between the groups. The regularity lemma is formulated as an asymptotical and existential statement. The graph we work with is definitely small, not comparable in size to those graphs to which one would in principle apply the regularity lemma. Thus our model can be viewed as a form of fitting, with allowance for error. We do not try to pretend that the assumptions of the regularity lemma apply to the case of large-scale cortical networks, but the idea underlying the regularity lemma, i.e. the probabilistic description of connections between and inside vertex groups, is exercised in order to find a good reconstruction. The proposed solution of the graph partitioning problem and its usage in edge prediction is achieved by using probabilistic methods, which allow finding solutions close to the optimum and satisfy the precision dictated by the practical applications.

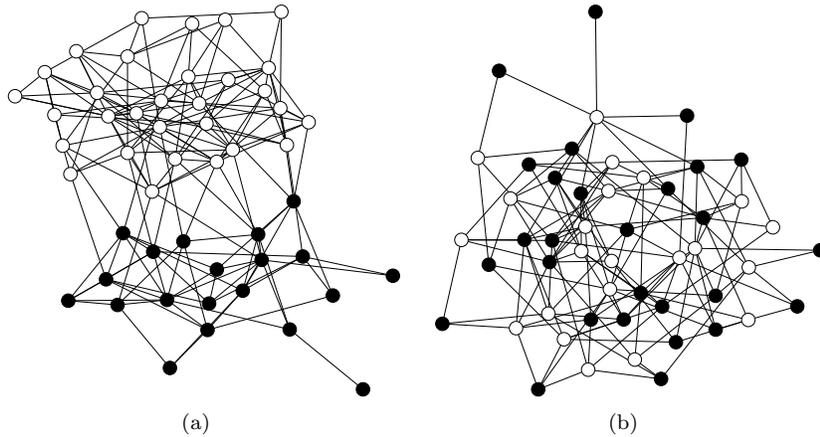


Fig. 1 Two graphs generated by the preference model. Black and white colours denote the groups the nodes belong to. Panel (a) shows a clustered graph where the connection probability between nodes within the same group is 0.2, while the connection probability between nodes in different groups is only 0.02. Panel (b) shows a bipartite graph where only nodes in different groups are allowed to connect with a probability of 0.2.

2.2 The preference model

This graph model starts from an empty graph with N vertices and it assigns every vertex to one of K distinct groups. The groups are denoted by integer numbers from 1 up to K . The generation process considers all pairs of vertices once, and it adds an edge between node v_1 and v_2 with probability p_{ij} if v_1 belongs to group i and v_2 belongs to group j . That is, the expected density of edges between group i and group j is exactly p_{ij} , and the existence of an edge between two vertices depends solely on the group affiliation of the vertices involved. Fig. 1 shows two possible graphs generated by this model. The one shown on Fig. 1(a) is a graph with clustered organization: vertices in similar groups tend to link together while rarely linking to vertices of the other group. The graph on Fig. 1(b) is a bipartite graph. The model allows the simultaneous appearance of these two basic patterns in a graph: $p_{ii} \approx 1$ results in the one seen on Fig. 1(a) and $p_{ij} \approx 1, i \neq j$ induces the one on Fig. 1(b). A method using similar ideas but designed for different applications was also described in a recent paper of Newman and Leicht [49].

The generalization of the model to directed graphs is straightforward: vertices of a directed graph will be assigned to an *incoming* and an *outgoing group* (in-group and out-group in short), and the probability of the existence of an edge between a vertex from out-group i and another vertex from in-group j is given by p_{ij} . The number of parameters in this model is $K^2 + 2N + 1$, since there are $2N$ parameters for the group affiliations of the vertices, K^2 parameters represent the elements of the preference matrix and the last pa-

parameter is K itself. The probabilities are usually arranged in a *probability matrix* \mathbf{P} for the sake of convenience. To emphasize the role of directionality, elements of the preference matrix in the directed case are sometimes denoted by $p_{i \rightarrow j}$ instead of p_{ij} . We also introduce the membership vectors $\mathbf{u} = [u_1, u_2, \dots, u_N]$ and $\mathbf{v} = [v_1, v_2, \dots, v_N]$, where u_i is the out-group and v_i is the in-group of vertex i . From now on, parameterizations of the model will be denoted by $\mathcal{M} = (K, \mathbf{u}, \mathbf{v}, \mathbf{P})$.

This model naturally gives rise to densely connected subnetworks with sparse connections between them by appropriately specifying the connection probabilities within and between groups. This is a characteristic property of cortical networks, and it is assumed that a good reconstruction of the network can be achieved by specifying vertex groups and connection probabilities appropriately. More precisely, given a graph $G(V, E)$ without multiple or loop edges, the reconstruction task is equivalent to specifying the number of groups, finding an appropriate assignment of vertices to groups and determining the elements of the probability matrix \mathbf{P} . The reconstructed graph then can be generated by the preference model, and new (previously unknown) connections can also be predicted by checking the probabilities of the uncertain edges in the fitted model. E.g., a crude reconstruction of the visuo-tactile network of the macaque monkey (see Section 3 for details about this dataset) would be a model with two groups (group 1 corresponding to the visual and group 2 to the tactile vertices in the network) and connection probabilities $p_{1 \rightarrow 1} = 0.385$, $p_{1 \rightarrow 2} = 0.059$, $p_{2 \rightarrow 1} = 0.035$ and $p_{2 \rightarrow 2} = 0.377$, based on the density of connections between the groups in the original network. The introduction of more vertex types results in a better reconstruction, and obviously the reconstruction is perfect when $N = K$ and \mathbf{P} is \mathbf{A} , the adjacency matrix of the graph. However, such a reconstruction is not able to predict unknown connections. We will discuss the problem of overfitting in Section 2.5.

2.3 Measuring the accuracy of reconstruction

Since the preference model is a probabilistic model, every possible graph with N vertices can theoretically be generated by almost any parameterization of the model, but of course some graphs are more likely to be generated by a specific parameterization than by others. Therefore, we measure the fitness of a particular parameterization $\mathcal{M} = (K, \mathbf{u}, \mathbf{v}, \mathbf{P})$ with respect to a given graph $G(V, E)$ by its likelihood, i.e. the probability of the event that the probabilistic model with parameters \mathcal{M} generates $G(V, E)$:

$$L(\mathcal{M}|G) = \prod_{(i,j) \in E} p_{u_i \rightarrow v_j} \prod_{\substack{(i,j) \notin E \\ i \neq j}} (1 - p_{u_i \rightarrow v_j}) \quad (1)$$

The restriction $i \neq j$ in the second product term corresponds to the nonexistence of loop edges (even if they exist, they are ignored). To avoid numerical errors when working with small probabilities, one can use the log-likelihood instead, for the log-likelihood attains its maximum at the same \mathcal{M} where the likelihood does:

$$\log L(\mathcal{M}|G) = \sum_{(i,j) \in E} \log p_{u_i \rightarrow v_j} + \sum_{\substack{(i,j) \notin E \\ i \neq j}} \log(1 - p_{u_i \rightarrow v_j}) \quad (2)$$

2.4 Fitting the preference model

Fitting a model to a given graph $G(V, E)$ is equivalent to the maximum likelihood estimation (MLE) of the parameters of the model with respect to the graph, i.e. choosing \mathcal{M} in a way that maximizes $\log L(\mathcal{M}|G)$. Since the number of possible group assignments is K^N (where K is the number of groups and $N = |V|$ is the number of vertices), which is exponential in N , direct maximization of $\log L(\mathcal{M}|G)$ by an exhaustive search is not feasible. An alternative, greedy approach is therefore suggested to maximize the likelihood.

2.4.1 Greedy optimization

Starting from an initial configuration $\mathcal{M}^{(0)} = (K, \mathbf{u}^{(0)}, \mathbf{v}^{(0)}, \mathbf{P}^{(0)})$, the greedy optimization will produce a finite sequence of model parameterizations $\mathcal{M}^{(0)}$, $\mathcal{M}^{(1)}$, $\mathcal{M}^{(2)}$, \dots satisfying $L(\mathcal{M}^{(k)}|G) \geq L(\mathcal{M}^{(k-1)}|G)$ for $k \geq 1$. First we note that the log-likelihood of an arbitrary configuration \mathcal{M} is composed of N local likelihood functions corresponding to the vertices:

$$\begin{aligned} \log L(\mathcal{M}|G) &= \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \log (A_{ij} p_{u_i \rightarrow v_j} + (1 - A_{ij}) (1 - p_{u_i \rightarrow v_j})) \\ &= \sum_{i=1}^N \log L_i(G|\mathcal{M}) \end{aligned} \quad (3)$$

where A_{ij} is 1 if there exists an edge from i to j and 0 otherwise. Let us assume first that K is given in advance. Starting from random initial membership vectors $\mathbf{u}^{(0)}$ and $\mathbf{v}^{(0)}$ of $\mathcal{M}^{(0)}$, we can estimate an arbitrary element $p_{i \rightarrow j}$ of the real underlying probability matrix \mathbf{P} by counting the number of edges that originate from out-group i and terminate in in-group j and divide it by the number of possible edges between out-group i and in-group j . The estimated probabilities are stored in $\mathbf{P}^{(0)}$. After that, we examine

the local likelihoods $L_i(G|\mathcal{M}^{(0)})$ for all vertices and choose the in- and out-groups of the vertices in a way that greedily maximizes their local likelihood, assuming that the group affiliations of all other vertices and the estimated probabilities remain unchanged. Formally, let $\mathbf{u}_{i=k}^{(0)}$ denote the vector obtained from $\mathbf{u}^{(0)}$ by replacing the i th element with k and similarly let $\mathbf{v}_{i=l}^{(0)}$ denote the vector obtained from $\mathbf{v}^{(0)}$ by replacing the i th element with l . Let $\mathcal{M}_{i,k,l}^{(0)} = (K, \mathbf{u}_{i=k}^{(0)}, \mathbf{v}_{i=l}^{(0)}, \mathbf{P}^{(0)})$, and for every vertex i , every out-group k and every in-group l , calculate $\log L_i(G|\mathcal{M}_{i,k,l}^{(0)})$. After that, put vertex i in out-group k and in-group l if that maximizes $\log L_i(G|\mathcal{M}_{i,k,l}^{(0)})$. Now calculate the next estimation of the probability matrix, $\mathbf{P}^{(1)}$, maximize the local log-likelihoods based on the new probability matrix and repeat these two alternating steps until $\mathbf{u}^{(k)} = \mathbf{u}^{(k-1)}$ and $\mathbf{v}^{(k)} = \mathbf{v}^{(k-1)}$.

2.4.2 Markov chain Monte Carlo sampling

The group assignments obtained by the greedy algorithm suffer from a minor flaw: they correspond only to a local maximum of the parameter space and not the global one. The local maximum means that no further improvement could be made by putting any single vertex in a different group while keeping the group affiliations of all other vertices intact. However, there is the possibility of improving the partition further by moving more than one vertex simultaneously. Another shortcoming of the algorithm is the danger of overfitting: partitions with high likelihood might perform poorly when one tries to predict connections, because they are too much fine-tuned to the graph being analyzed. Therefore we also consider employing Markov chain Monte Carlo (MCMC) sampling methods [3] on the parameter space. (An alternative MCMC-based data mining method on networks is presented in [13], but while that method infers hierarchical structures in networks, our algorithm is concerned with the discovery of densely connected subgraphs and bipartite structures; see Fig. 1(a) and Fig. 1(b), respectively).

Generally, MCMC methods are a class of algorithms for sampling from a probability distribution that is hard to be sampled from directly. These methods generate a Markov chain whose equilibrium distribution is equivalent to the distribution we are trying to sample from. In our case, the samples are parameterizations of the preference model, and the distribution we are sampling from is the following:

$$\mathbf{P}(\mathcal{M} = \mathcal{M}_0) = \frac{L(\mathcal{M}_0|G)}{\int_{\mathbf{S}_K} L(\mathcal{M}'|G) d\mathcal{M}'} \quad (4)$$

where \mathbf{S}_K is the space of all possible parameterizations of the probability model for a given K . Informally, the probability of drawing \mathcal{M} as a sample should be proportional to its likelihood of generating $G(V, E)$, for instance, if \mathcal{M}_1 generates our network with a probability of 0.5 and \mathcal{M}_2 generates it with a probability of 0.25, \mathcal{M}_1 should be drawn twice as frequently as \mathcal{M}_2 .

The generic framework of the MCMC method we use is laid down in the Metropolis-Hastings algorithm [25]. The only requirement of the algorithm is that a function proportional to the density function (that is, $\mathbf{P}(\mathcal{M} = \mathcal{M}_0)$ in (4)) can be calculated. Note that $\mathbf{P}(\mathcal{M} = \mathcal{M}_0) \propto L(\mathcal{M}_0|G)$, since the denominator in (4) is constant. Starting from an arbitrary random parameterization $\mathcal{M}^{(0)}$, MCMC methods propose a new parameterization \mathcal{M}' based on the previous parameterization $\mathcal{M}^{(t)}$ using a proposal density function $Q(\mathcal{M}'|\mathcal{M}^{(t)})$. If the proposal density function is symmetric ($Q(\mathcal{M}'|\mathcal{M}^{(t)}) = Q(\mathcal{M}^{(t)}|\mathcal{M}')$), the probability of accepting the proposed parameterization is $\min(1, L(\mathcal{M}'|G)/L(\mathcal{M}^{(t)}|G))$. When the proposal is accepted, it becomes the next state in the Markov chain ($\mathcal{M}^{(t+1)} = \mathcal{M}'$), otherwise the current state is retained ($\mathcal{M}^{(t+1)} = \mathcal{M}^{(t)}$).

MCMC sampling can only approximate the target distribution, since there is a residual effect depending on the starting position of the Markov chain. Therefore, the sampling consists of two phases. In the first phase (called *burn-in*), the algorithm is run for many iterations until the residual effect diminishes. The second phase is the actual sampling. The burn-in phase must be run long enough so that the residual effects of the starting position become negligible.

A desirable property of a Markov chain in a MCMC method is *rapid mixing*. A Markov chain is said to mix rapidly if its mixing time grows at most polynomially fast in the logarithm of the number of possible states in the chain. Mixing time refers to a given formalization of the following idea: how many steps do we have to take in the Markov chain to be sure that the distribution of states after these steps is close enough to the stationary distribution of the chain? Given a guaranteed short mixing time, one can safely decide to stop the burn-in phase and start the actual sampling after the number of steps taken exceeded the mixing time of the chain.

Several definitions exist for the mixing time of a Markov chain (for an overview, see [43]). To illustrate the concept, we refer to a particular variant called *total variation distance mixing time*, which is defined as follows:

Definition 1 (Total variation distance mixing time). Let S denote the set of states of a Markov chain \mathcal{C} , let $A \subseteq S$ be an arbitrary nonempty subset of the state set, let $\pi(A)$ be the probability of A in the stationary distribution of \mathcal{C} , and $\pi_t(A)$ be the probability of A in the distribution observed after step t . The total variation distance mixing time of \mathcal{C} is the smallest t such that $|\pi_t(A) - \pi(A)| \leq 1/4$ for all $A \subseteq S$ and all initial states.

However, many practical problems have resisted rigorous theoretical analysis. This applies also to the method presented here, mostly due to the fact

that the state transition matrix of the Markov chain (and therefore its stationary distribution) is a complicated function of the adjacency matrix of the network and the number of vertex groups, and no closed form description exists for either. In these cases, a common approach to decide on the length of the burn-in phase is based on the acceptance rate, which is the fraction of state proposals accepted during the last m steps. Sampling is started when the acceptance rate drops below a given threshold (a typical choice is 20% or 0.2). Local maxima are avoided by accepting parameterization proposals with a certain probability even when they have a lower likelihood than the last one, but being biased at the same time towards partitions with high likelihoods. In the case of multiple local maxima with approximately the same likelihood, MCMC sampling tends to oscillate between those local maxima. By taking a large sample from the equilibrium distribution, one can approximate the probability of vertex i being in out-group k and in-group l and extract the common features of all local maxima (vertices that tend to stay in the same groups despite randomly walking around in the parameter space).

The only thing left to clarify before employing MCMC sampling on fitting the preference model is the definition of an appropriate symmetric proposal density function. We note that the number of groups K is constant and the probability matrix \mathbf{P} can be approximated by the edge densities for a given out- and in-group assignment, leaving us with only $2N$ parameters that have to be determined. We take advantage of the fact that the conditional distribution of each parameter (assuming the others are known) can be calculated exactly as follows:

$$\mathbf{P}(u_i = k) = \frac{L_i(G|\mathcal{M}_{i,k,*})}{\sum_{l=1}^K L_i(G|\mathcal{M}_{i,l,*})} \quad (5a)$$

$$\mathbf{P}(v_i = k) = \frac{L_i(G|\mathcal{M}_{i,*,k})}{\sum_{l=1}^K L_i(G|\mathcal{M}_{i,*,l})} \quad (5b)$$

where $\mathcal{M}_{i,k,*} = (K, \mathbf{u}_{i=k}, \mathbf{v}, \mathbf{P})$ and $\mathcal{M}_{i,*,k} = (K, \mathbf{u}, \mathbf{v}_{i=k}, \mathbf{P})$. Since the conditional distribution of each parameter is known, Gibbs sampling [24] can be used. The Gibbs sampling alters a single variable of the parameter vector in each step according to its conditional distribution, given all other parameters. It can be shown that the proposal distribution defined this way is symmetric if the variable being modified is picked randomly according to a uniform distribution. In practice, it is sufficient to cycle through the variables in a predefined order as long as the Markov chain can access all states under this ordering. To speed up the burn-in process, one can apply the greedy optimization method described in Section 2.4.1 and revert to the MCMC sampling when the algorithm reached the first local maximum.

2.5 Choosing the number of groups

As mentioned earlier in Section 2.2, the key parameter that controls the balance between accurate reconstruction and meaningful prediction is the number of vertex groups used in the preference model. A very small number of groups yields an inaccurate reconstruction and most likely meaningless predictions. Increasing the number of groups gradually improves the accuracy of reconstruction, attaining perfection when the number of groups is equal to the number of vertices, but in this case no new edges are predicted. This is the classical problem of overfitting: by increasing the number of groups, the ability of the model to generalize beyond the original data diminishes. Therefore, the goal is to select the number of groups in a way that achieves good reconstruction while still allowing the model to predict connections by assigning a high probability to vertex pairs where an uncertain connection is suspected.

We tried multiple approaches to infer the appropriate number of groups in the networks we studied. The exact results will be discussed in Section 2.6.2 and Section 3; here we only outline the basic ideas. We will make use of the eigenvalues of the Laplacian matrix of the graph, the singular value decomposition (SVD) of the adjacency matrix and the Akaike information criterion [2].

Given an undirected graph $G(V, E)$ without loops and multiple edges, its Laplacian matrix is defined as $\mathcal{L} = \mathbf{D} - \mathbf{A}$, where \mathbf{A} is the adjacency matrix and \mathbf{D} is a diagonal matrix composed of the degrees of the vertices. A basic property of the Laplacian matrix is that its smallest eigenvalue is zero, and its multiplicity is equal to the number of connected components of the graph. The number of eigenvalues close to zero is frequently used for determining the number of dense subgraphs (communities, clusters) in the graph and, based on similar reasoning, this could be a good estimate of the number of groups that have to be used in the preference model; however, we cannot use $\mathbf{D} - \mathbf{A}$ directly, since this form of the Laplacian is defined only for undirected graphs.

An extension of the Laplacian to directed graphs was introduced in [12]. This involves calculating the Perron vector ϕ of the transition probability matrix \mathbf{P} of the graph. The transition probability matrix \mathbf{P} is derived from the adjacency matrix by normalizing the row sums to be 1. The Perron vector ϕ is a unique (up to scaling) left eigenvector of \mathbf{P} satisfying $\phi\mathbf{P} = \phi$. The existence of this vector is guaranteed by the Perron-Frobenius theorem. There is no closed-form solution for ϕ , but it is easy to calculate in polynomial time numerically. The directed Laplacian is then defined as:

$$\mathcal{L} = \mathbf{I} - \frac{\Phi^{1/2}\mathbf{P}\Phi^{-1/2} + \Phi^{-1/2}\mathbf{P}^*\Phi^{1/2}}{2} \quad (6)$$

where \mathbf{P}^* is the conjugate transpose of \mathbf{P} and Φ is a diagonal matrix composed of the elements of ϕ , assuming that $\sum_{i=1}^n \phi_i = 1$. The properties emphasized above for the undirected Laplacian hold for the directed Laplacian as well.

The singular value decomposition of an $m \times n$ matrix \mathbf{M} is a factorization process that produces an $m \times m$ and an $n \times n$ unitary matrix (\mathbf{U} and \mathbf{V} , respectively) and an $m \times n$ matrix Σ with non-negative numbers on the diagonal and zeros off the diagonal in a way that $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^*$. The diagonal of Σ contains the singular values, while the columns of \mathbf{U} and \mathbf{V} are the left and right singular vectors, respectively. Plotting the singular values on a scree plot (sorted from large to small) is a good visual cue to determining the number of groups in the model: the number of groups can simply be assigned according to the number of large singular values. It is noteworthy that one can approximate the original matrix \mathbf{M} by setting all singular values other than the l largest to zero and disregarding the appropriate rows of \mathbf{U} and \mathbf{V} that correspond to the zeroed singular values. The remaining parts of \mathbf{U} and \mathbf{V} can serve as an input for a k -means clustering algorithm in an l -dimensional space, and the results of the clustering yield a good candidate of an initial position of the greedy optimization process of the preference matrix. In practice, however, performing a complete SVD is less efficient than optimization from a random initial position.

The Akaike information criterion (AIC) [2] is a measure of the goodness of fit of a statistical model (the preference model in our case). It is an unbiased estimator of the Kullback-Leibler divergence [39], and it is an operational way of determining the appropriate trade-off between the complexity of a model and its predictive power. AIC is calculated as $2k - 2 \log L$, where k is the number of parameters in the model and L is the likelihood. In the preference model, $k = K^2 + 2N + 1$. The suggested number of groups can be determined by fitting the model with various numbers of groups and choosing the one that minimizes the Akaike information criterion.

The AIC can also be used to detect situations when the network being studied is in fact completely random, and therefore its appropriate description is simply an Erdős-Rényi random graph model instead of the preference model. This is done by estimating the probability parameter p of the Erdős-Rényi model from the edge density of the network and then calculating the log-likelihood of the network according to the Erdős-Rényi model. Given a directed network with n vertices and m edges, the maximum likelihood estimator of p is $\frac{m}{n(n-1)}$, resulting in a log-likelihood of $m \log p + (n^2 - n - m) \log(1 - p)$ (assuming that there are no loop edges). The baseline AIC corresponding to the Erdős-Rényi model is then $2 - 2(m \log p + (n^2 - n - m) \log(1 - p))$, since the model has only a single parameter. If the network being studied is completely random, the AIC corresponding to the case of two groups will be larger than the baseline AIC of the Erdős-Rényi model, for we introduced more parameters without actually improving the likelihood. On the other hand, networks possessing a structure that can be described by

the preference model will show significant improvement in the log-likelihood compared to the pure random case, resulting in a lower AIC.

2.6 Performance measurements

To demonstrate the validity of the fitting algorithms presented above, we conducted several benchmarks on computer-generated test graphs. First, we generated graphs according to the preference model, ran the fitting algorithm on the graphs by supplying the appropriate number of groups beforehand and then compared the known and the estimated parameters of the model. These benchmarks were performed in order to test the validity of the fitting algorithm and to assess the quality of the results obtained. Next, we ran the fitting algorithms without specifying the number of groups to show that the Akaike information criterion is suitable for determining the right value of k .

2.6.1 Fitting the model with given number of groups

This benchmark proceeded as follows: graphs with 128 vertices were generated according to the preference model using 4 in- and out-types. The type distribution was uniform, so there were 32 vertices of each type on average. The preference matrix was chosen as follows: each element p_{ij} was set to one of two predefined values p_1 and p_2 with probability 0.5. p_1 and p_2 was varied between 0 and 1 with a step size 0.05. For each (p_1, p_2) combination, we generated 50 graph instances using the preference model. Values of the quality functions (described below) were averaged over these instances and the results were plotted as a function of p_1 and p_2 . We used only two probabilities because the results can then be visualized on a heat map or a 2.5D plot.

To assess the fitness of the fitted model, we had to define some quality functions that compare the fitted parameters to the original (expected) ones. First we note that the number of groups and the probability matrix do not have to be compared, since the former is fixed and the latter one is calculated from the group assignments, so errors in the elements of the probability matrices are simply due to errors in the group assignments. Therefore, only the group assignments matter. The following quality functions were defined:

Normalized mutual information of the confusion matrix. This measure was suggested by Fred and Jain [22] and later applied to community detection in graphs by Danon et al. [15]. The measure is based on the *confusion matrix* $\mathbf{C} = [c_{ij}]$ of the expected and observed group assignments. c_{ij} is the number of vertices that are in group i in the original and group j in the fitted model. The confusion matrix can be calculated separately for in- and out-groups, but they can safely be added together to obtain

a single confusion matrix and then a single quality measure, which is the normalized mutual information of the confusion matrix:

$$I(\mathbf{C}) = -2 \frac{\sum_{i=1}^k \sum_{j=1}^k c_{ij} \log \frac{c_{ij} c_{**}}{c_{i*} c_{*j}}}{\sum_{i=1}^k \left(c_{i*} \log \frac{c_{i*}}{c_{**}} + c_{*i} \log \frac{c_{*i}}{c_{**}} \right)} \quad (7)$$

where c_{i*} is the sum of the i -th row, c_{*j} is the sum of the j -th column of the confusion matrix. c_{**} is the sum of c_{ij} for all i, j . It is assumed that $0 \log 0 = 0$. When the fitted group assignment is completely identical to the expected one (apart from rearrangement of group indices), $I(\mathbf{C})$ attains its maximum at 1. $I(\mathbf{C}) = 0$ if the two group assignments are independent. Danon et al. [15] argue that this measure is in general stricter than most other quality measures proposed so far. For instance, a completely random assignment of groups still has an expected success ratio of 0.25 for 4 groups (since each pair is consistent with probability 1/4). In this case, the normalized mutual information is close to zero, which is a more intuitive description of what happened than a success ratio of 0.25. See the paper of Danon et al. [15] for a list of other measures they considered.

Likelihood ratio. This measure is simply the ratio of the likelihoods of the original and the fitted parameterizations, given the generated graph.

The likelihood ratios and the mutual information indices are plotted on Figure 2. As expected, the mutual information index is low when $p_1 \approx p_2$. This is no surprise, since $p_1 \approx p_2$ implies that the actual difference between different vertex types diminish: they all behave similarly, and the random fluctuations at this network size render them practically indistinguishable. The overall performance of the algorithm is satisfactory in the case of $p_1 \ll p_2$ and $p_1 \gg p_2$, with success ratios and mutual information indices larger than 0.9 in all cases. In cases when $p_1 \approx p_2$, the likelihood ratio is greater than 1, which indicates that the fitted model parameterization is more likely than the original one. This phenomenon is an exemplar of overfitting: apparent structure is detected by the algorithm where no structure exists at all if we use too many groups.

2.6.2 Fitting the model without a predefined number of groups

In Section 2.5, we described three different methods for estimating the number of groups one should use for a given network when fitting the preference model. Two of these methods requires some human intervention, since one had to choose a threshold manually for the eigenvalues of the Laplacian matrix or for the singular values of the adjacency matrix.

We investigated the eigenvalues of the directed Laplacian matrix first. After some experiments on graphs generated according to the preference model, it became obvious that the number of eigenvalues of the Laplacian close to

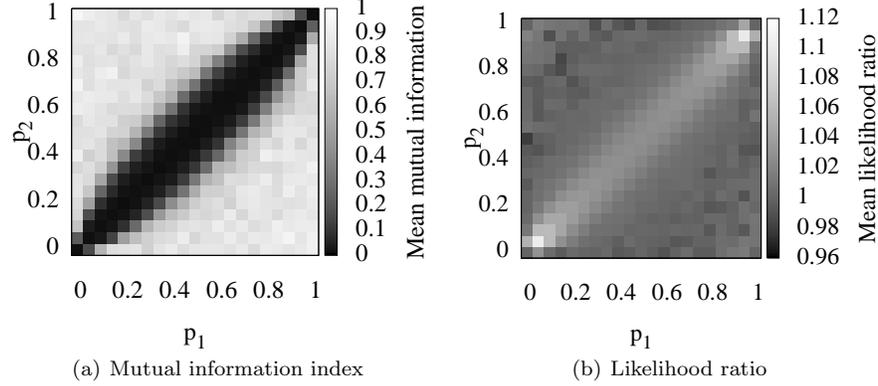


Fig. 2 Mean likelihood ratios of the fitted parameterizations to the expected ones (left) and mean normalized mutual information conveyed by the confusion matrices (right) as a function of p_1 and p_2 .

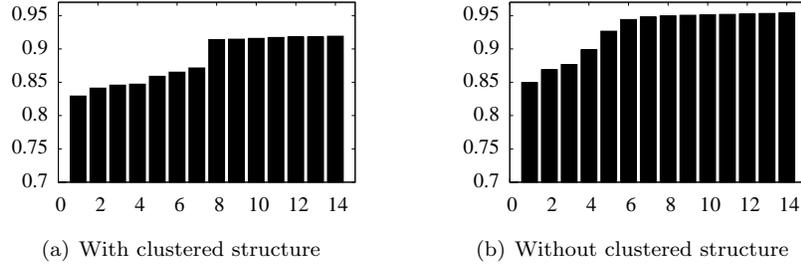


Fig. 3 The 15 smallest *nonzero* eigenvalues of the Laplacian matrix for graphs generated by the preference model with 8 groups, either with or without a strong clustered structure (left and right panel, respectively)

zero correlate to the number of groups only if the vertex groups coincide with densely connected subgraphs. In other words, p_{ii} must be large and p_{ij} for $i \neq j$ must be small. This is illustrated on Figure 3. The left panel shows the case when $p_{ij} = 0.2 + 0.6\delta(i, j)$ (the graph is clustered) and the right panel shows the case when p_{ij} is 0.2 or 0.8 with 1/2 probability. There is indeed a relatively large jump after the eighth eigenvalue for the former case, but the transition is smooth for the latter. Therefore, the eigenvalues of the directed Laplacian matrix were excluded from further investigations.

In the case of SVD analysis, one has to count the large singular values. “Large” is definitely a subjective term, therefore a scree plot of the singular values is often used as a visual aid. The scree plot is simply a bar graph of the singular values sorted in their decreasing order. The plot usually looks like the side of a mountain with some debris at the bottom: the singular

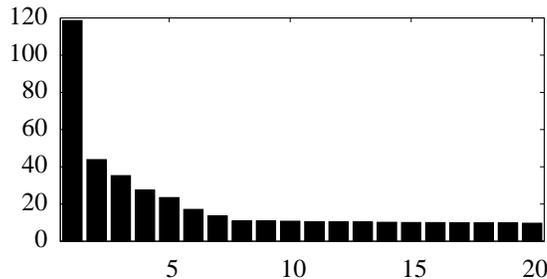


Fig. 4 The largest 20 singular values of the adjacency matrix of a graph generated by the preference model with 8 groups

values decrease rapidly at first, but there is an elbow where the steepness of the slope decreases abruptly, and the plot is almost linear from there on (see Figure 4 for an illustration). The number of singular values to the left of the elbow is the number of groups we will choose. To allow for automated testing, we implemented a simple method to decide on the place of the elbow. The approach we used is practically equivalent to the method of Zhu and Ghodsi [69]. It is based on the assumption that the values to the left and right of the elbow behave as independent samples drawn from a distribution family with different parameters. The algorithm first chooses a distribution family (this will be the Gaussian distribution in our case), then considers all possible elbow positions and calculates the maximum likelihood estimation of the distribution parameters based on the samples to the left and right side of the elbow. Finally, the algorithm chooses the position where the likelihood was maximal. Assuming Gaussian distributions on both sides, the estimates of the mean and variance are as follows:

$$\begin{aligned} \tilde{\mu}_1 &= \frac{\sum_{i=1}^q x_i}{q} & \tilde{\mu}_2 &= \frac{\sum_{i=q+1}^n x_i}{n-q} \\ \tilde{\sigma}^2 &= \frac{\sum_{i=1}^q (x_i - \mu_1)^2 + \sum_{i=q+1}^n (x_i - \mu_2)^2}{n-2} \end{aligned} \quad (8)$$

where x_i is the i -th element in the scree plot (sorted in decreasing order), n is the number of elements (which coincides with the number of vertices) and q is the number of elements standing to the left of the elbow. Note that the means of the Gaussian distributions are estimated separately, but the variance is common. Zhu and Ghodsi [69] argue that allowing different variances makes the model too flexible. The common variance is calculated by taking into account that the first q elements are compared to μ_1 and the remaining ones are compared to μ_2 . See the paper of Zhu and Ghodsi [69] for a more detailed description of the method.

In this benchmark, 100 networks were generated with 128 vertices each. Elements of the preference matrix were chosen to be p_1 or p_2 with equal probability, as in Section 2.6.1 before, but the case of $p_1 \approx p_2$ was avoided by constraining p_1 to be above 0.6 and p_2 to be below 0.4. The number of groups was varied between 2 and 8 according to a uniform distribution. The number of groups in the fitted model was estimated by the SVD and the AIC methods, the best AIC was chosen by trying all possible group counts between 2 and 10. The AIC method proved to be superior to the SVD method: the estimation was perfect in 79% of the cases. The number of groups was underestimated by 1 group in 14, 2 groups in 3 and 3 groups in 2 cases. There were 2 overestimations by 1 group as well, resulting in a mean squared error of 0.46 groups. On the other hand, the SVD method made severe mistakes at times; in fact, only 7% of its estimations matched the prior expectations, all other cases were overestimations, sometimes by 7 or 8 groups. This is due to the unsupervised choice of the elbow in the scree plot. It is assumed that better results can be achieved by making the choice manually, therefore the conclusion is that the SVD-based estimation should be handled with care and the AIC method is preferred when one would like to choose the number of groups automatically.

2.7 Handling uncertain connections

Despite being concerned about predicting unknown connections in a network where some parts are uncertain, we only discussed fitting the preference model to a graph where all connections were known and all uncertain connections were assumed to be nonexistent. As a refinement of the model, we can include our *a priori* assumption about the probability of the event that a particular, presently uncharted connection exists in the network. Let us denote by $b_{i \rightarrow j}$ our degree of belief in the existence of an edge going from vertex i to j . We write $b_{i \rightarrow j} = 1$ if we are fully convinced that the edge actually exists and $b_{i \rightarrow j} = 0$ for edges that are known to be nonexistent. Intermediary values of $b_{i \rightarrow j}$ can be thought about as probabilities, e.g., $b_{i \rightarrow j} = 0.3$ means that the probability of an edge from vertex i to j is 0.3. (Note that $b_{i \rightarrow j}$ acts as a generalization of A_{ij} : $b_{i \rightarrow j}$ is 1 if and only if we are convinced that $A_{ij} = 1$, $b_{i \rightarrow j}$ is 0 if and only if we are convinced that $A_{ij} = 0$. Uncertain connections result in $0 < b_{i \rightarrow j} < 1$). In this sense, not only our model but the graph being fitted is also probabilistic, and we are trying to find the model whose expected likelihood with respect to the whole ensemble of possible graphs parameterized by the degrees of belief is maximal. All the optimization methods described earlier also work in this case, only the likelihood and the log-likelihood functions have to be adjusted:

$$L(\mathcal{M}|G) = \prod_{i=1}^N \prod_{j=1, j \neq i}^N (b_{i \rightarrow j} p_{u_i \rightarrow v_j} + (1 - b_{i \rightarrow j})(1 - p_{u_i \rightarrow v_j})) \quad (9a)$$

$$\log L(\mathcal{M}|G) = \sum_{i=1}^N \sum_{j=1, j \neq i}^N \log (b_{i \rightarrow j} p_{u_i \rightarrow v_j} + (1 - b_{i \rightarrow j})(1 - p_{u_i \rightarrow v_j})) \quad (9b)$$

The elements of the optimal probability matrix \mathbf{P} can then be thought about as the posterior probabilities of the edges in the network. An edge whose prior probability is significantly lower than its posterior probability is then likely to exist, while connection candidates with significantly higher prior than posterior probabilities are likely to be nonexistent.

3 Results and discussion

In this section, we will present our results on application of the preference model to the prediction of unknown connections in the visual and sensorimotor cortex of the primate (macaque monkey) brain.

The dataset we are concerned with in this section is a graph model of the visuo-tactile cortex of the macaque monkey brain. Connectivity data was retrieved from the CoCoMac database [37] and it is identical to the dataset previously published in [45]. The whole network contains 45 vertices and 463 directed links among them. The existence of connections included in the network were confirmed experimentally, while connections missing from the network were either explicitly checked for and found to be nonexistent, or never checked experimentally. To illustrate the uncertainty in the dataset being analyzed, we note that 1157 out of the 1980 possible connections were uncertain (never checked experimentally) and only 360 were known to be absent.

The network consists of two dense subnetworks corresponding to the visual and the sensorimotor cortex (30 and 15 vertices, respectively). The visual cortex can also be subdivided to the so-called dorsal and ventral parts using a community detection algorithm based on random walks [40]. Most of the uncertain connection candidates are heteromodal (originating in the visual and terminating in the sensorimotor cluster, or the opposite), and it is assumed that the vast majority of possible heteromodal connections are indeed nonexistent. The basic properties of these networks are shown in Table 1, while the adjacency matrix of the visuo-tactile network is depicted on Fig. 5. Note that since the visual and sensorimotor cortices are subnetworks of the visuo-tactile networks, their adjacency matrices are the upper-left 30×30 and lower-right 15×15 submatrices of the adjacency matrix of the visuo-tactile cortex. In order to compare our results with previous reconstruction attempts that were only concerned with the visual cortex [21, 30], we will present results based on the visual subnetwork as well as the whole visuo-tactile cortex.

Table 1 Basic properties of the original networks

	Visual	Sensorimotor	Visuo-tactile
Vertices	30	15	45
Known connections (edges)	335	85	463
Known nonexistent connections	310	0	360
Unknown connections	225	125	1157
Density	0.385	0.404	0.233
Density (excl. unknowns)	0.519	1.000	0.548
Diameter	3	3	5
Average path length	1.6632	1.767	2.149
Reciprocity	0.850	0.888	0.815

Data analysis was performed using the open source `igraph` network analysis library [14] (<http://cneurocv.s.rmki.kfki.hu/igraph>).

3.1 Rapid mixing of the MCMC process

First, we illustrate the rapid convergence of the MCMC process to the equilibrium distribution. This property is crucial, since a chain with short mixing time tends to get close to its equilibrium distribution quickly, thus ensuring that a short burn-in period is sufficient. Since a rigorous proof of the mixing time of the Markov chain designed for fitting the preference model is well beyond the scope of this chapter, and we are mostly concerned with its applicability to the visual and visuo-tactile networks, we check the fast convergence of the method by plotting the log-likelihood of the states of the Markov chain and the acceptance rate as the fitting progresses from an arbitrary random starting position. Fig. 6 illustrates that the chain mixes rapidly, reaching its equilibrium distribution in roughly n^2 steps, where n is the number of vertices in the network. This satisfies the criterion of rapid mixing, since the number of possible states in the Markov chain is k^{2n} (k is the number of groups), therefore the mixing time is polynomial in the logarithm of the number of states. The number of groups was fixed at 7 for the visual and 10 for the visuo-tactile cortex, these choices will be explained later. Unknown edges were treated as nonexistent. Fig. 6 suggests that one can start sampling from the Markov chain after roughly n^2 steps or after the acceptance rate drops below 0.2.

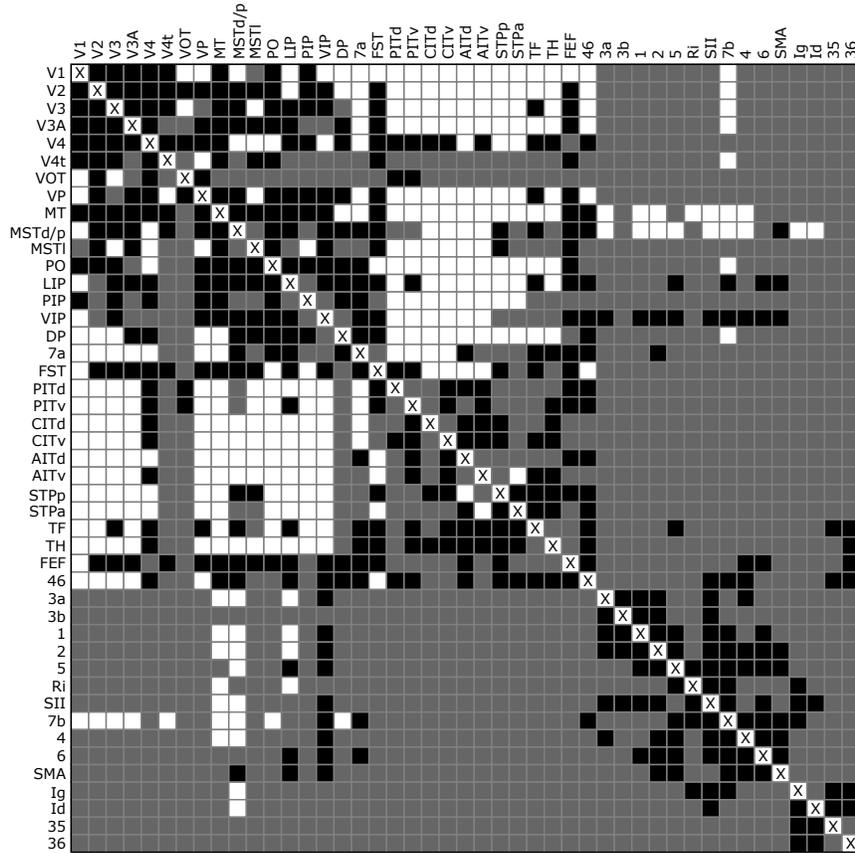


Fig. 5 Adjacency matrix of the visuo-tactile cortex dataset. Black cells denote known existing connections, white cells denote known non-existent connections. Gray cells are connections not confirmed or confuted experimentally. The upper left 30×30 submatrix is the adjacency matrix of the visual cortex, the lower right 15×15 submatrix describes the sensorimotor cortex.

3.2 Methodological comparison with other prediction approaches

Our method allows prediction of nonreciprocal connections, and the network data is not symmetrised for the sake of computational and methodological tractability, in contrast to [21]. Furthermore, we only use the connectional data for prediction, no other anatomical facts were taken into account. An approach where additional neuroanatomical facts were used as predictonal input is described in [21]. Jouve et al. [30] use a specific property of the visual cortex: the existence of indirect connections of length 2 between areas are

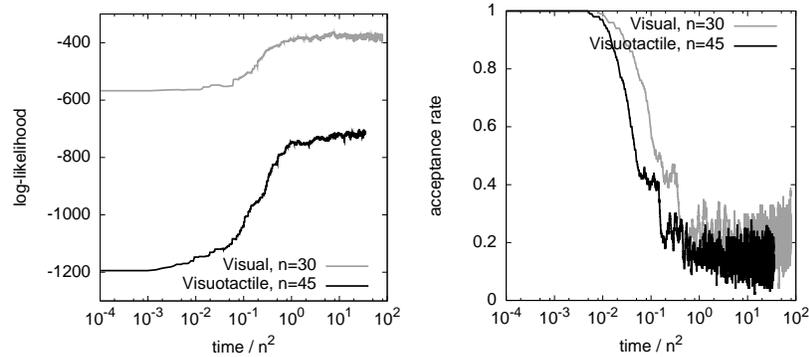


Fig. 6 Log-likelihood of the states of the Markov chain (left) and acceptance rates in a window of 100 samples (right) for the visual and visuo-tactile cortices, normalised by n^2 , on a logarithmic time scale

presumed to support the existence of a direct connection. This property need not hold for other large cortical structures, especially when investigating the interplay of different cortices (e.g., the visual and the sensorimotor cortices). In fact, this assumption is difficult to prove or disprove due to the poor knowledge of connection structure in other parts of the cortex. The problem is even more pronounced in the case of heteromodal connections, thus other guiding principles had to be sought.

3.3 Visual cortex

Since the visual cortex is a part of the visuo-tactile cortex, the adjacency graph of the visual cortex can be found in Fig. 5 as the upper left 30×30 submatrix. It is noteworthy that most of the unknown connections are adjacent to the areas VOT and V4t, and the subgraph consisting of the vertices PITd, PITv, CITd, CITv, AITd, AITv, STPp and STPa (all belonging to the ventral class) is also mostly unknown. Based on the connection density of the visual cortex (assuming unknown connections to be nonexistent), the probability of the existence of a connection classified as unknown was set to 0.385. These degrees of belief were taken into account in the likelihood function as described in Sect. 2.7. The search for the optimal configuration started from a random initial position, first improved by a greedy initial phase, then followed by MCMC sampling after reaching the first local maximum. The sampling process was terminated when at least 10^6 samples were taken from the chain. The sample with the best likelihood became the final result.

The optimal number of groups in the preference model was determined by studying the eigenvalues of the Laplacian and the singular values of the ad-

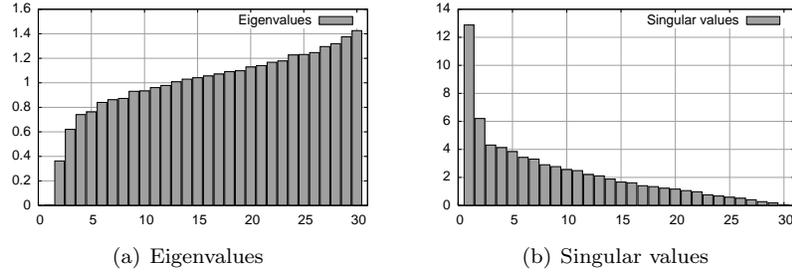


Fig. 7 Eigenvalues of the Laplacian and singular values of the adjacency matrix of the visual cortical graph

adjacency matrix (unknown connections were treated as nonexistent) as well as the Akaike information criterion of the obtained partitions at various group numbers from 2 to 15. Partitions having more than 15 groups do not seem feasible, since in these cases, at least one of the groups will contain only one vertex. The eigenvalues and the singular values are shown on Fig. 7. A visual inspection suggests using only two groups (which is congruent with the anatomical fact that the visual cortex is composed of two major pathways, namely the dorsal and the ventral stream), but the minimal AIC value was achieved using 7 groups (see Table 2). Since two groups are intuitively insufficient for an accurate reconstruction, we decided to use 7 groups in the rest of the analysis. This is further supported by the mediocre success rate of the model with only two groups. Success rates were calculated as follows: for every possible threshold τ between 0 and 1 (with a granularity of 0.01), the percentage of known edges that had a predicted probability greater than τ was calculated. The final threshold used for calculating the success rate was chosen to be the one that produced the highest ratio of correctly predicted known edges. τ fluctuated around 0.5 in all cases. As it was expected based on our reasoning outlined in Sect. 2.5, the success rate increased steadily as we increased the number of groups, but the divergence of τ from 0.5 after having more than 7 groups is likely to be a precursor of overfitting.

The fitted model with 7 groups provided probabilities for the 225 unknown connections, 137 of them were above the optimal threshold $\tau = 0.5$. The ratio of predicted edges approximately matches the density of the visual cortex when we exclude the unknown connections from the density calculation (see Table 1). However, if we wanted the ratio of predicted connections to match the density of known connections in the visual cortex, we would have to increase τ to 0.654, predicting only 81 connections. This ratio matches the one reported in [21], although the connection matrix in [21] included an additional area in the analysis. The predicted adjacency matrix with $\tau=0.654$ is shown on Fig. 8 and its basic descriptive graph measures are to be found in Table 3.

Table 2 Likelihoods, AIC values and success rates in the visual cortex

K	Log-likelihood	AIC	τ	Success rate
2	-481.608	1091.216	0.50	80.7%
3	-440.280	1018.560	0.50	82.4%
4	-413.027	978.055	0.50	84.2%
5	-394.664	959.328	0.50	85.1%
6	-378.271	948.543	0.50	87.4%
7	-363.146	944.292	0.50	87.6%
8	-353.071	954.143	0.50	88.5%
9	-340.886	963.773	0.47	89.6%
10	-331.626	983.253	0.43	90.2%
11	-319.771	1001.543	0.49	90.7%
12	-307.766	1023.532	0.48	91.5%
13	-300.657	1059.315	0.48	91.9%
14	-297.540	1107.081	0.46	92.0%
15	-288.615	1147.231	0.49	92.4%

Table 3 Basic properties of the predicted networks

	Predicted visual	Predicted visuo-tactile
Vertices	30	45
Edges	358	757
Density	0.412	0.382
Diameter	3	4
Average path length	1.478	1.833
Reciprocity	0.517	0.645
Reciprocity of predicted connections	0.674	0.824

Besides the overall success rate, we also calculated the ratio of correctly predicted 1's and 0's (R_1 and R_0 , respectively) for the case of $K = 7$. With $\tau = 0.5$, 84.8% of known 0's and 93.1% of known 1's were predicted correctly ($R_0 = 0.848$, $R_1 = 0.931$). The geometric mean ($\sqrt{R_0 R_1}$) was 0.888, which dropped to 0.8245 when raising τ to 0.654 ($R_0 = 0.925$, $R_1 = 0.734$), thus a higher τ seems to be better at reconstructing non-existing connections.

We compared our results to earlier studies [21,30]. Comparisons were based on the percentage of matching predictions. Since both studies took a slightly different sets of areas into consideration, we did not take into account those areas that were not present in any of the matrices.

The predictions of Jouve et al. [30] are based solely on topological features of the network model of the visual cortex, similarly to the method presented here. The agreement between the two predicted matrices is moderate: 61.6% of the predictions match for $\tau = 0.5$ and only 47% for $\tau = 0.654$. Most of the disagreements involved areas V4t (28), VOT (22), FEF (17) and DP (16). Area MSTd was joined together with MSTp in our study (resulting in the vertex denoted by MSTd/p), therefore neither MSTd nor MSTd/p was taken into account. We note that the matrix used in the present paper incorporated

	V1	V2	V3	V3A	V4	V4t	VOT	VP	MT	MSTd/p	MSTl	PO	LIP	PIP	VIP	DP	7a	FST	PITd	PITv	CITd	CITv	AITd	AITv	STPp	STPa	TF	TH	FEF	46	
V1	0	1	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
V2	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0
V3	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0
V3A	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0
V4	1	1	1	1	0	1	1	0	1	0	0	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
V4t	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
VOT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	
VP	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0
MT	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0
MSTd/p	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	1	0	1	0	1	1	
MSTl	0	1	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PO	1	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
LIP	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	1	0	1	1	
PIP	1	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VIP	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	1	0	1	0	1	
DP	1	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7a	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	
FST	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	1	0	1	0	1	1	
PITd	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1
PITv	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1
CITd	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	1
CITv	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1
AITd	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	1	0	1
AITv	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	1	0	1
STPp	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1
STPa	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	1	1	0	1
TF	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
TH	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1
FEF	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	1	0	1	0	0	1
46	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0

Fig. 8 The predicted adjacency matrix of the visual cortex with 7 vertex groups. White cells denote confirmed existing and absent connections. Dark gray cells denote mismatches between the known and the predicted connectivity. Light gray cells denote predictions for unknown connections.

the results of anatomical experiments that could not have been included in the matrix in [30], therefore the moderate match between the two matrices can be explained by the differences in the initial dataset, see Table 4 for the number of mismatches involving each area. Since the prediction method of Jouve et al. was not concerned with reconstructing the entire network (predictions were made only on unknown connections), no comparison could be made based on the success rates of the two methods.

The predictions published by Costa et al. [21] are based on several topological (e.g., node degree, clustering coefficient) and spatial features (e.g., area sizes, local density of the areas in the 3D space, based on their known positions in the cortex). In this sense, the reconstruction method based on the preference model is simpler, for it depends solely on the connection ma-

Table 4 Number of mismatches in the predicted matrix, grouped by areas

	Known connections	Jouve et al. [30]	Costa et al. [21]
V1	5	0	0
V2	5	1	0
V3	6	0	0
V3A	1	4	1
V4	<u>21</u>	0	0
V4t	2	<u>28</u>	<u>15</u>
VOT	7	<u>22</u>	<u>18</u>
VP	10	0	–
MT	3	2	0
MSTd/p	5	–	–
MSTl	7	7	0
PO	10	3	1
LIP	8	3	3
PIP	5	13	4
VIP	4	6	3
DP	<u>12</u>	<u>16</u>	8
7a	9	10	7
FST	11	8	1
PITd	6	6	2
PITv	4	7	6
CITd	2	9	3
CITv	3	8	5
AITd	7	2	1
AITv	4	8	3
STPp	11	9	8
STPa	6	12	3
TF	<u>13</u>	8	<u>11</u>
TH	6	8	2
FEF	8	<u>17</u>	3
46	<u>15</u>	7	<u>9</u>

In the second column, the known connections of the original matrix are compared to our predictions. In the last two columns, only the *unknown* (predicted) connections are compared to the *unknown* connections of our dataset. The 4 largest number of mismatches in each column are underlined.

trix. We also note that Costa et al. inferred the topological features from a symmetrized connectivity matrix, thus their predicted matrix is also completely symmetric, while our method produced a matrix where only 67.4% of the connections (67.4% of the predicted, previously unknown connections) were reciprocal. The ratios of correctly predicted 1’s and 0’s in the visual cortex reported by Costa et al. were slightly worse ($R_0 = 244/350 = 0.697$, $R_1 = 207/295 = 0.701$, $\sqrt{R_0 R_1} = 0.699$, loop connections excluded). Note that the comparison can not be fully accurate because of the slightly different set of areas used in the analysis (MIP and MDP were present only in [21], whereas MSTd/p and VP were present only in the matrix used in this study). 69.8% of the predictions presented here matched the predictions of [21], and

all predicted edges with a probability larger than 0.8 were predicted in [21] as well.

One may note that in spite of the improvement of the reconstruction as compared to the previous studies [21, 30], there is still a relatively high number of mismatches on Fig. 8. The distribution of mismatches in the adjacency matrix can be suggestive of the methodological shortcomings and the state of knowledge in the investigated network. It appears that most of the mismatches are to be found within the two major visual clusters, the dorsal and ventral visual subsystems, where connectional densities are higher than in the lower left and upper right quadrants of the matrix representing the inter-cluster connections. Interestingly, most of the mismatches affected either the input or output patterns of areas V4 and 46, and to a lesser degree of TF and FEF in the intercluster regions. These areas are central nodes in the visual cortical network, establishing connections between different clusters. In fact, the inclusion of the sensorimotor cortex improved the reconstruction (see Section 3.4). It is also noteworthy that relatively few mismatches/violations occurred in case of the lower order areas (listed in the upper left corner of the matrix). This is an important point as low-level areas establish connections mostly within their cluster and the connections of these areas are relatively well explored. These observations indicate the dependence of reconstruction quality on the actual knowledge of the network.

To summarize without going into the details, we conclude that our reconstruction is biologically realistic and reflects our understanding of the organization of the visual cortical connectivity.

3.4 *Visuo-tactile cortex*

The network model of the visuo-tactile cortex is an extension of the visual cortex, obtained by adding the 15 areas of the sensorimotor cortex and their respective connections. Connections going between a visual and a sensorimotor area are called heteromodal connections. The density of the sensorimotor cortex is slightly higher than that of the visual cortex. Based on the connection densities, the probability of the existence of an unknown connection was assumed to be 0.385 inside the visual cortex and 0.404 inside the sensorimotor cortex. Unknown heteromodal connections were assumed to exist with probability 0.1. Note that the vast majority of heteromodal connections is unknown. There was no confirmed nonexisting sensorimotor connection indicated in the data set. The adjacency matrix is shown on Fig. 5. The optimal configuration was found by combining the greedy optimization with the MCMC method, similarly as above.

The number of groups in the preference model was determined again by the Akaike information criterion. The eigenvalues of the Laplacian and the singular values of the adjacency matrix suggested 5 groups, which is again in

Table 5 Likelihoods, AIC values and success rates in the visuo-tactile cortex

K	Log-likelihood	AIC	τ	Success rate
5	-814.956	1859.913	0.42	83.6%
6	-783.935	1819.871	0.23	84.4%
7	-756.352	1790.705	0.46	84.8%
8	-736.163	1780.327	0.37	86.1%
9	-718.422	1778.844	0.43	86.4%
10	-697.078	1774.156	0.49	87.3%
11	-683.335	1788.671	0.46	89.3%
12	-684.105	1836.210	0.46	89.3%
13	-665.337	1848.674	0.47	89.4%
14	-653.755	1879.510	0.48	89.4%
15	-652.173	1934.347	0.40	90.1%

concordance with anatomical considerations, but as shown above, 5 groups were insufficient to reproduce only the visual cortex (part of the visuo-tactile cortex). Log-likelihoods, AIC values and success rates are shown in Table 5, from 5 to 15 groups. The optimal number of groups with the lowest AIC was 10.

The fitted model with 10 groups predicted 225 connections with $\tau = 0.47$ out of the 1157 unknown ones ($R_0 = 0.883$, $R_1 = 0.892$, $\sqrt{R_0 R_1} = 0.887$). This is 19.4% of the unknown connections and it roughly matches the overall density of the visuo-tactile cortex (23.3%). However, only 5 heteromodal connections (all originating from LIP) were predicted apart from the known existing ones. This is due to the fact that very little is known about the heteromodal connections, and the algorithm cannot generalize beyond them with higher confidence. We also note that the posterior probability of many heteromodal connections in this case stayed at 0.1, the same as their prior probability. Taking into account that even a significant difference between the prior and the posterior probabilities of the heteromodal connections may not reach the threshold of 0.49, we decided to use different thresholds for non-heteromodal and heteromodal connections (τ_1 and τ_2 , respectively). τ_1 was left at 0.49, while τ_2 was lowered to 0.137, the average *a posteriori* probabilities of the unknown heteromodal connections. This new configuration yielded $R_0 = 0.831$, $R_1 = 0.927$, $\sqrt{R_0 R_1} = 0.877$ and 132 predicted heteromodal connections, related mainly to areas LIP, VIP, DP, 7a, FST, TF, FEF and 46 in the visual cortex. It is noteworthy that four of these areas (46, 7a, LIP and VIP) were classified as structural overlaps between the two subnetworks in the fuzzy community analysis of Nepusz et al. [47]. Anatomical considerations also support the bridge-like role of these areas between the cortices. It was previously suggested in the literature that area VIP should be split into two areas (VIPm and VIPp), establishing stronger connections with visual or sensorimotor areas, respectively [42]. VIP and LIP are involved with hand and eye coordination, respectively, requiring a combined input of visual and tactile signals. Area 46 is a part of the dorsolateral prefrontal cortex, and it

Table 6 Group affiliations of the areas in the visuo-tactile cortex

Out-group 1	V1, VOT, MSTl
Out-group 2	V2, V3, V3A, V4t, VP, MT, PO, PIP
Out-group 3	V4
Out-group 4	MSTd/p, FST, FEF
Out-group 5	LIP, VIP
Out-group 6	DP, 7a
Out-group 7	PITd, PITv, CITd, CITv, AITd, AITv, STPp, STPa, TH
Out-group 8	TF, 46
Out-group 9	3a, 1, 2, 5, SII, 7b, 4, 6, SMA
Out-group 10	3b, Ri, Ig, Id, 35, 36
In-group 1	V1, PIP
In-group 2	V2, V3, V3A, V4t, VP, MT, PO
In-group 3	35, 36
In-group 4	V4, FST, FEF
In-group 5	VIP
In-group 6	MSTd/p, MSTl, LIP, DP, 7a
In-group 7	PITd, PITv, CITd, CITv, AITd, AITv, STPp, STPa, TH
In-group 8	VOT, TF, 46
In-group 9	3a, 1, 2, 5, SII, 7b, 4, 6, SMA
In-group 10	3b, Ri, Ig, Id

does not have functions related to low-level sensory information processing. Being a higher level (supramodal) area, it integrates visual, tactile and other information. Area 7a integrates visual, tactile and proprioceptive signals. Finally, areas TF and FEF are also high level structures integrating widespread cortical information (e.g., [20]).

The predicted connectivity matrix is shown on Fig. 9, the basic graph measures are depicted in Table 3. To show the subtle differences between predicted connections, the exact probabilities are shown on Fig. 10, encoded in the background colour of the matrix cells (white indicating zero probability and black indicating 1). The latter figure shows the prediction in its full detail, especially in the sensorimotor cortex where the predicted clique-like subgraph reveals its internal structure more precisely. The group affiliations of the individual vertices are shown in Table 6.

We also examined the ratios of correctly predicted known 0's and 1's with respect to pure visual and pure sensorimotor connections. As expected, the algorithm performed better in the visual cortex, which is more thoroughly charted than the sensorimotor cortex. The calculated ratios were $R_0 = 0.865$, $R_1 = 0.902$, $\sqrt{R_0 R_1} = 0.882$ for the visual cortex. Since the sensorimotor cortex contained no known non-existing connections, R_1 could not be calculated for it. All known connections in the sensorimotor cortex were predicted correctly ($R_0 = 1$), however, this is due to the lack of information on nonexisting connections in the sensorimotor cortex. The ratios of the visual cortex were similar to the ones obtained when analyzing the visual cortex alone.

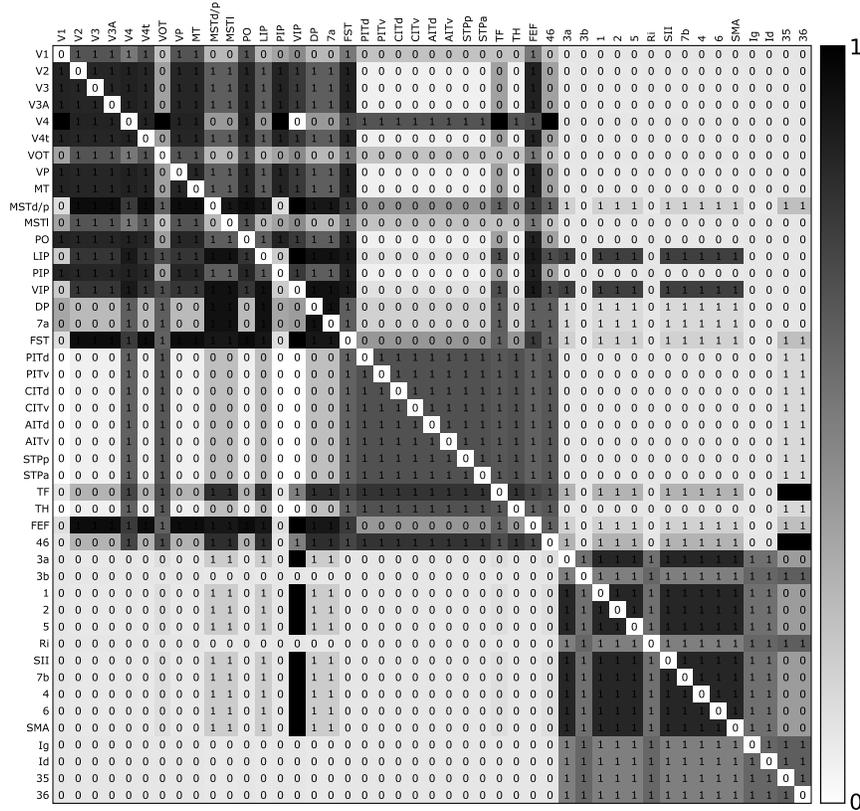


Fig. 10 Probability of the connections in the visuo-tactile cortex with 10 vertex groups, $\tau_1 = 0.49$ and $\tau_2 = 0.137$. Probabilities are denoted by colours, with white corresponding to 0 and black corresponding to 1. The predicted adjacency matrix is shown in the matrix cells.

(see Fig. 8 and the upper left part of Fig. 9). This was not a simple consequence of increasing the number of clusters from 7 to 10, but the corollary of the additional information about the connections that visual areas form with the sensorimotor cortex. This contextual information may also give guidelines for understanding the mechanisms of heteromodal interactions.

3.5 Major structural changes after reconstruction

An interesting feature of the reconstructed network is the complete clique induced by the following vertices: PITd, PITv, CITd, CITv, AITd, AITv, STPp, STPa, TF and TH. We note that this region was mostly uncharted

in the original matrix, and there were only three confirmed nonexisting connections ($STP_p \rightarrow AIT_d$ and $STP_a \leftrightarrow AIT_v$). This clique also appeared in the earlier analysis of Jouve et al. [30] and a similar but smaller clique also emerged in our predictions of the visual cortex. Notably, a similar tendency could also be observed in the analysis of Costa et al. [21], as only a few connections were missing among the areas mentioned above to form a clique. Our assumption is that our prediction based on solely the visual cortex is a more accurate approximation of the true connectional pattern of these areas.

Finally, we compared the community structure of the original and the predicted connectivity matrix in order to obtain further support for the validity of our predictions. We argue that a valid predicted connectivity matrix should not only obtain a high ratio of correctly predicted 1's and 0's, but also keep the community structure of the network mostly intact. In order to take the directionality of the edges into account, we employed the community detection method of Latapy & Pons [40]. This method is based on the idea that given a significant community structure in the network, short random walks tend to stay within the same community. We tried random walks with length ranging from 3 to 7. The quality of the obtained partitions was assessed by the modularity function Q [48]. Best results were achieved with random walks of length 7 for the original and 4 for the predicted cortex ($Q = 0.339$ and $Q = 0.301$, respectively). Table 7 shows the detected communities of the original and the predicted visuo-tactile cortex.

Table 7 Community structure of the original and the predicted visuo-tactile cortex

Original visuo-tactile cortex ($Q = 0.339$)	
Community 1	V1, V2, V3, V3A, V4t, VP, MT, MSTd/p, MSTl, PO, LIP, PIP, VIP, DP, 7a, FST, FEF, 46
Community 2	V4, VOT, PITd, PITv, CITd, CITv, AITd, AITv, STPp, STPa, TF, TH
Community 3	3a, 3b, 1, 2, 5, Ri, SII, 7b, 4, 6, SMA
Community 4	Ig, Id, 35, 36
Predicted visuo-tactile cortex ($Q = 0.379$)	
Community 1	V1, V2, V3, V3A, V4t, VP, MT, MSTd/p, MSTl, PO, LIP, PIP, VIP, DP, 7a, FST, FEF, 46
Community 2	V4, VOT, PITd, PITv, CITd, CITv, AITd, AITv, STPp, STPa, TF, TH
Community 3	3a, 3b, 1, 2, 5, Ri, SII, 7b, 4, 6, SMA, Ig, Id, 35, 36

Community 1 corresponds approximately to the dorsal stream of the visual cortex while community 2 contains the areas of the ventral stream in both cases. Community 3 and community 4 form the sensorimotor cortex. These groups were joined together to form a single community of the sensorimotor cortex in the case of the predicted matrix. Since there is no known anatomical meaning of community 4, the predicted community structure represents our understanding of the sensorimotor cortex more accurately than the original one. Apart from this mismatch, we can conclude that the major structure of

the original matrix was preserved during the reconstruction process, but the reconstructed network seems less modular than the original, as indicated by the decrease in the modularity of the partitions.

Other notable differences were the smaller diameter and shorter average path length (Table 1, 3). Considering segregation and integration as major cortical functions [66], these findings together predict a more efficient cortical information processing as it can be deduced on the basis of our present state of knowledge about this network.

4 Conclusion

From a biological point of view, the stochastic representation of the cortical network described in the present study demonstrates the high level of uncertainty of our knowledge about the connectivity of the ventral visual cluster (community 2) as well as the sensorimotor cortex (communities 3 and 4). Similarly to our findings, both previous studies [21,30] predicted numerous connections within the ventral visual cluster, making it almost a large, fully connected clique. A similar observation was made here in the sensorimotor cortex. We should note that in contrast to the dorsal visual cluster (including the majority of areas in community 1) there is a massive lack of information regarding the non-existence of the connections in the ventral and sensorimotor clusters (see also [20]). Considering the deterministically imperfect knowledge of the cortical, and probably most of the real-life complex networks, these findings point out the significance of information about known nonexistent connections. Interestingly, similar conclusion was drawn by Kötter and Stephan [38], who investigated the role of components in cortical information processing by defining their network participation indices. This is an important point when considering that neurobiologists focus on the existence of connections between the cortical areas and often ignore the absence of them while mapping cortical connections in their experiments. In fact, this point is in agreement with our expectations considering that the cortical network architecture is shaped by evolutionary and functional constraints.

Considering individual cortical areas, our predictions resulted in a relatively large number of mismatches of the connections of some mid- (e.g., V4 and DP) and high level areas (e.g., TF and 46) when compared to the original data. The connections of high level areas, which form the connections between the different cortical clusters, are hard to predict in individual sub-networks where the connections between these areas and the neglected part of the network is necessarily missing from the data. The largest number of mismatches occurred in the case of V4. This area is a functionally heterogeneous, complex structure, which maybe divided to sub-regions, and forms a bridge between the dorsal and ventral clusters of the visual cortex [47].

This could at least partly explain the large number of mismatches occurred in the reconstruction process. The assumption outlined here is further supported by the fact that the number of mismatches involving areas V4, DP, TF and 46 decreased after taking into account the sensorimotor cluster and the heteromodal connections. On the other hand, V4t and VOT are examples of structures with largely unexplored connectivity. The existing data suggest that these areas exhibit roughly similar pattern of connectivity to their topographical neighbor area V4, which has much better explored connectivity. The large differences observed in the predicted connectivity of V4t and VOT in the three studies (including ours as well as Costa et al. [21] and Jouve et al. [30]) is most probably due to the large uncertainty in the connectional data of these two areas. Interestingly, a more careful examination of Fig. 8 suggest that the prediction of the connections of these three areas, especially that of V4 and VOT tended to be somewhat complementary in the visual cortex. This was less evident in the larger visuo-tactile network (Fig. 9). Based on these observations it is suggested that the mismatches of V4-connections occurred during the reconstruction as a consequence of the optimization process.

From the methodological point of view it is important to note that our approach is *not* a community structure detection. There are many sophisticated methods which reliably detect communities within networks [17, 40, 48, 53], yet the information about the community structure is not sufficient for an accurate network reconstruction. One of the main advantages of our method is a higher accuracy in reconstruction of charted connections (known existent or nonexistent), compared to earlier methods [21, 30]. It is important to emphasize that our method allows the prediction of nonreciprocal connections. The presented approach is general and applicable to other networks not necessarily related to cortical structure. Its generality is reflected in the fact that we do not use information about the underlying spatial structure, as these data may be unavailable or unintelligible for some networks, nor we assume that there is a given property of paths related to the clustering coefficient in the network. Our method is based on an exact goal function that we optimize. The existence of a goal function allows for a comparison of different solutions which may be difficult to carry out in the absence of a quantitative measure. The drawbacks of our approach are related to the necessity of knowledge regarding existing and confirmed nonexisting connections. This necessity is reflected in the prediction of heteromodal connections, as the number of confirmed or refuted heteromodal connections is surprisingly small, see Fig. 5. One has to note that the method is computationally expensive, which may be a severe limiting factor in case of large networks. Finally, inspired by the Szemerédi regularity lemma, the present study exemplifies the usefulness of a theoretical approach in analyzing real world data.

Acknowledgments

László Négyessy acknowledges the support of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

1. Achard, S., Bullmore, E.: Efficiency and cost of economical brain functional networks. *PLoS Comp Biol* **3**(2), e17 (2007)
2. Akaike, H.: A new look at the statistical model identification. *IEEE Trans Automat Contr* **19**(6), 716–723 (1974)
3. Andrieu, C., de Freitas, N., Doucet, A., Jordan, M.I.: An introduction to MCMC for machine learning. *Mach Learn* **50**, 5–43 (2003)
4. Angelucci, A., Bullier, J.: Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J Physiol* **97**(2-3), 141–54 (2003)
5. Arbib, M.A., Érdi, P.: Précis of Neural organization: structure, function and dynamics. *Behav Brain Sci* **23**(4), 513–33 (2000). Discussion: pp. 533-71
6. Bienenstock, E.: On the dimensionality of cortical graphs. *J Physiol* **90**(3-4), 251–6 (1996)
7. Braitenberg, V., Schüz, A.: *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer (1998)
8. Bressler, S.L., Tognoli, E.: Operational principles of neurocognitive networks. *Int J Psychophysiol* **60**(2), 139–48 (2006)
9. Buzás, P., Kovács, K., Ferecskó, A.S., Budd, J.M., Eysel, U.T., Kisvárday, Z.F.: Model-based analysis of excitatory lateral connections in the visual cortex. *J Comp Neurol* **499**(6), 861–81 (2006)
10. Buzsáki, G., Geisler, C., Henze, D.A., Wang, X.J.: Interneuron diversity series: Circuit complexity and axon wiring economy of cortical interneurons. *Trends Neurosci* **27**(4), 186–93 (2004)
11. Calvin, W.H.: Cortical columns, modules, and Hebbian cell assemblies. In: M.A. Arbib (ed.) *Handbook of Brain Theory and Neural Networks*, pp. 269–272. Bradford Books/MIT Press (1995)
12. Chung, F.R.K.: Laplacians and the Cheeger inequality for directed graphs. *Ann Combinator* **9**, 1–19 (2005)
13. Clauset, A., Moore, C., Newman, M.E.J.: Structural inference of hierarchies in networks. In: *Proceedings of the 23rd International Conference on Machine Learning (ICML)* (2006)
14. Csárdi, G., Nepusz, T.: The igraph software package for complex network research. *InterJournal Complex Systems* p. 1695 (2006)
15. Danon, L., Duch, J., Diaz-Guilera, A., Arenas, A.: Comparing community structure identification. *J Stat Mech* p. P09008 (2005)
16. DeFelipe, J., Alonso-Nanclares, L., Arellano, J.I.: Microstructure of the neocortex: comparative aspects. *J Neurocytol* **31**(3-5), 299–316 (2002)
17. van Dongen, S.: A stochastic uncoupling process for graphs. Tech. Rep. INS-R0011, National Research Institute for Mathematics and Computer Science in the Netherlands (2000)
18. Douglas, R.J., Martin, K.A.: Mapping the matrix: the ways of neocortex. *Neuron* **56**(2), 226–38 (2007)
19. Edelman, G.M.: Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron* **10**(2), 115–25 (1993)

20. Felleman, D.J., Van Essen, D.C.: Distributed hierarchical processing in the primate cerebral cortex. *Cerebr Cortex* **1**(1), 1–47 (1991)
21. da Fontoura Costa, L., Kaiser, M., Hilgetag, C.C.: Predicting the connectivity of primate cortical networks from topological and spatial node properties. *BMC Syst Biol* **1**(16) (2007)
22. Fred, A.L.N., Jain, A.K.: Robust data clustering. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 128–133. IEEE (2003). DOI 10.1109/CVPR.2003.1211462
23. Freeman, W.J.: Mesoscopic neurodynamics: from neuron to brain. *J Physiol* **94**(5-6), 303–22 (2000)
24. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* **6**, 721–741 (1984)
25. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**(1), 97–109 (1970). DOI 10.2307/2334940
26. Hilgetag, C.C., Burns, G.A., O’Neill, M.A., Scannell, J.W., Young, M.P.: Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philos Trans R Soc Lond B Biol Sci* **355**(1393), 91–110 (2000)
27. Hilgetag, C.C., Grant, S.: Uniformity, specificity and variability of corticocortical connectivity. *Philos Trans R Soc Lond B Biol Sci* **355**(1393), 7–20 (2000)
28. Hilgetag, C.C., O’Neill, M.A., Young, M.P.: Hierarchical organization of macaque and cat cortical sensory systems explored with a novel network processor. *Philos Trans R Soc Lond B Biol Sci* **355**(1393), 71–89 (2000)
29. Ioannides, A.A.: Dynamic functional connectivity. *Curr Opin Neurobiol* **17**, 161–70 (2007)
30. Jouve, B., Rosenstiehl, P., Imbert, M.: A mathematical approach to the connectivity between the cortical visual areas of the macaque monkey. *Cerebr Cortex* **8**(1), 28–39 (1998)
31. Kaas, J.H.: Organizing principles of sensory representations. *Novartis Foundation symposium* **228**, 188–198 (2000)
32. Kaiser, M.: Brain architecture: a design for natural computation. *Philos Transact A Math Phys Eng Sci* **365**(1861), 3033–45 (2007)
33. Kaiser, M., Hilgetag, C.C.: Modelling the development of cortical systems networks. *Neurocomputing* **58-60**, 297–302 (2004)
34. Kaiser, M., Hilgetag, C.C.: Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comp Biol* **2**(7), e95 (2006)
35. Kisvárday, Z.F., Ferecskó, A.S., Kovács, K., Buzás, P., M., B.J., Eysel, U.T.: One axon-multiple functions: specificity of lateral inhibitory connections by large basket cells. *J Neurocytol* **31**(3-5), 255–64 (2002)
36. Komlós, J., Simonovits, M.: Szemerédi’s regularity lemma and its applications in graph theory. In: D. Miklós, V.T. Sós, T. Szőnyi (eds.) *Combinatorics, Paul Erdős is Eighty*, Vol. 2, pp. 295–352. János Bolyai Mathematical Society, Budapest (1996)
37. Kötter, R.: Online retrieval, processing, and visualization of primate connectivity data from the cocomac database. *Neuroinformatics* **2**, 127–144 (2004)
38. Kötter, R., Stephan, K.E.: Network participation indices: characterizing component roles for information processing in neural networks. *Neural Network* **16**(9), 1261–75 (2003)
39. Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann Math Stat* **22**, 79–86 (1951)
40. Latapy, M., Pons, P.: Computing communities in large networks using random walks. *J Graph Algorithm Appl* **10**(2), 191–218 (2006)
41. Lee, L., Friston, K., Horwitz, B.: Large-scale neural models and dynamic causal modelling. *NeuroImage* **30**, 1243–54 (2006)

42. Lewis, J.W., Van Essen, D.C.: Corticocortical connections of visual, sensorimotor and multimodal processing areas in the parietal lobe of the macaque monkey. *J Comp Neurol* **428**, 112–137 (2000)
43. Montenegro, R., Tetali, P.: Mathematical aspects of mixing times in Markov chains. *Foundations and Trends in Theoretical Computer Science* **1**(3), 237–354 (2006). DOI 10.1561/0400000003
44. Mountcastle, V.B.: The columnar organization of the neocortex. *Brain* **120**(4), 701–22 (1997)
45. Négyessy, L., Nepusz, T., Kocsis, L., Bazsó, F.: Prediction of the main cortical areas and connections involved in the tactile function of the visual cortex by network analysis. *Eur J Neurosci* **23**(2), 1919–1930 (2006)
46. Négyessy, L., Nepusz, T., Zalányi, L., Bazsó, F.: Convergence and divergence are mostly reciprocated properties of the connections in the network of cortical areas. *Proc Roy Soc B* **275**(1649), 2403–2410 (2008). DOI 10.1098/rspb.2008.0629
47. Nepusz, T., Petróczy, A., Négyessy, L., Bazsó, F.: Fuzzy communities and the concept of bridgeness in complex networks. *Phys Rev E* **77**, 016,107 (2008)
48. Newman, M.E.J.: Modularity and community structure in networks. *Proc Natl Acad Sci Unit States Am* **103**(23), 8577–8582 (2006)
49. Newman, M.E.J., Leicht, E.A.: Mixture models and exploratory analysis in networks. *Proc Natl Acad Sci Unit States Am* **104**(23), 9564–9569 (2007). DOI 10.1073/pnas.0610537104
50. Passingham, R.E., Stephan, K.E., Kötter, R.: The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci* **3**(8), 606–16 (2002)
51. Petroni, F., Panzeri, S., Hilgetag, C.C., Kötter, R., Young, M.P.: Simultaneity of responses in a hierarchical visual network. *NeuroReport* **12**(12), 2753–9 (2001)
52. <http://www.ncbi.nlm.nih.gov/PubMed/>
53. Reichardt, J., Bornholdt, S.: Detecting fuzzy community structures in complex networks with a Potts model. *Phys Rev Lett* **93**, 218,701 (2004)
54. Rockland, K.S.: Non-uniformity of extrinsic connections and columnar organization. *J Neurocytol* **31**(3-5), 247–53 (2002)
55. Rockland, K.S., N., I.: Some thoughts on cortical minicolumns. *Exp Brain Res* **158**(3), 256–77 (2004)
56. Roe, A.W.: Modular complexity of area V2 in the macaque monkey. In: C. Collins, J. Kaas (eds.) *The Primate Visual System*, pp. 109–138. CRC Press (2003)
57. Roe, A.W., Ts’o, D.Y.: Specificity of color connectivity between primate V1 and V2. *J Neurophysiol* **82**(5), 2719–30 (1999)
58. Sporns, O.: Graph theory methods for the analysis of neural connectivity patterns. In: R. Kötter (ed.) *Neuroscience Databases. A Practical Guide*. Klüwer (2002)
59. Sporns, O., Chialvo, D.R., Kaiser, M., Hilgetag, C.C.: Organization, development and function of complex brain networks. *Trends Cognit Sci* **8**(9), 418–25 (2004)
60. Sporns, O., Honey, C.J., Kötter, R.: Identification and classification of hubs in brain networks. *PLoS ONE* **2**(10), e1049 (2007)
61. Stettler, D.D., Das, A., Bennett, J., Gilbert, C.D.: Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron* **36**(4), 739–50 (2002)
62. Szemerédi, E.: Regular partitions of graphs. In: E. du Centre National de la Recherche Scientifique (CNRS) (ed.) *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976, pp. 399–401. Paris (1978)*
63. Szentágothai, J.: The “module-concept” in cerebral cortex architecture. *Brain Res* **95**(2-3), 475–96 (1975)
64. Szentágothai, J.: Specificity versus (quasi-) randomness revisited. *Acta Morphologica Hungarica* **38**, 159–167 (1990)
65. Tao, T.: Szemerédi’s regularity lemma revisited. *Contributions to Discrete Mathematics* **1**, 8–28 (2006)
66. Tononi, G., Edelman, G.M., Sporns, O.: Complexity and coherency: integrating information in the brain. *Trends Cognit Sci* **2**, 474–84 (1998)

67. Van Essen, D.C., Dierker, D.L.: Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* **56**, 209–25 (2007)
68. Young, M.P.: Objective analysis of the topological organization of the primate cortical visual system. *Nature* **358**(6382), 152–5 (1992)
69. Zhu, M., Ghodsi, A.: Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics and Data Analysis* **51**(2), 918–930 (2006). DOI 10.1016/j.csda.2005.09.010

Index

- Akaike information criterion, 12
- bridge vertex, 33
- cerebral cortex, 2
- CoCoMac, 3, 19
- cognitive functions, 3
- connectional similarity, 3
- cortical areas, 3
 - dorsal visual, 33
 - sensorimotor, 33
 - ventral visual, 33
- cortical column, 2
- directedness, 3
- experimental data, 3
- Gibbs sampling, 11
- graph model
 - preference, 4
 - stochastic, 4
- graph partitioning, 5
- greedy optimization, 8
- integration
 - neural, 2
- Kullback-Leibler divergence, 13
- likelihood, 7
- log-likelihood, *see* likelihood
- macaque monkey, 3
- Markov chain Monte Carlo method, 9
- matrix
 - adjacency, 12
 - directed Laplacian, 12
 - Laplacian, 12
 - maximum likelihood estimation, 8
 - membership vector, 7
 - Metropolis-Hastings algorithm, 10
- networks
 - stochastic reconstruction, 3
 - stochastic representation, 33
- neural connections
 - heteromodal, 33
 - indirect, 21
 - long distance, 2
 - nonexistent, 34
 - nonreciprocal, 34
- neuron doctrine, 2
- neuronal diversity, 2
- optimization
 - goal function, 34
 - stochastic, 34
- overfitting, 9, 12
- Perron vector, 12
- Perron-Frobenius theorem, 12
- quasi-randomness, 2
- reciprocity, 3
- segregation
 - neural, 2
- singular value decomposition, 12
- synapse, 2
- Szemerédi
 - Endre Szemerédi, 5
 - regularity lemma, 5
- uncertainty, 3