**Conditional Random Fields for Single Cell Targeting of Neuronal ensembles**

Luis Carrillo-Reid\*, Shuting Han\*, Tony Jebara, Rafael Yuste

**Highlights <85**

- CRFs depict neuronal ensembles representing features of visual stimuli**70**

- CRFs allows identification of core neurons from each neuronal ensemble**70**

- Core neurons evoke pattern completion of optogenetically imprinted ensembles**76**

- CRFs capture changes in network dynamics induced by circuit reconfiguration**75**

**Summary 139<150**

The prediction of different states from many variables with mutual dependencies is fundamental in various applications including natural language, computer vision and bioinformatics. Such problems can be approached by structured prediction methods that combine graphical models and classification algorithms. Conditional random fields (CRFs) represent a widely used probabilistic method for structured prediction. However, CRFs application to infer the functional connectivity of biological neural networks remains unexplored. We used CRFs and graph theory in population calcium imaging from primary visual cortex (V1) of awake behaving mice to identify neuronal ensembles and predict visual stimuli. Finally, using simultaneous two-photon imaging and optogenetics we show that our approach can be used to identify core neurons from neuronal ensembles imprinted *in vivo*. Our method provides a powerful tool for targeting individual neurons that can influence overall network activity.

**Introduction**

The coordinated firing of neuronal populations is considered to be the substrate of sensory, behavioral and cognitive functions. These coactive neuronal groups, defined as neuronal ensembles, are assumed to generate complex circuit functions that cannot be achieved by single neurons ([Miller et al., 2014](#_ENREF_28)). Recent advances in two-photon calcium imaging and two-photon optogenetics, have made possible the recording of simultaneous activity from large ensembles of neurons while manipulating population activity with single cell resolution in awake behaving animals ([Carrillo-Reid et al., 2016](#_ENREF_10)). However, how the functional connectivity in cortical microcircuits relate to their function has been difficult to elucidate because it requires the identification of physiologically relevant neurons that can be targeted during close-loop optogenetic experiments, allowing the manipulation of learned behavioral tasks.

Graph theory has been applied to model the structural and functional organization of the brain ([Bullmore and Sporns, 2009](#_ENREF_7)). However, graphs are usually constructed with nodes representing brain regions ([He et al., 2007](#_ENREF_19)), and edges representing information flow ([Iturria-Medina et al., 2008](#_ENREF_21)). For functional analysis, many studies have constructed graphs with data from fMRI, EEG and electrode arrays, taking brain regions ([Achard and Bullmore, 2007](#_ENREF_1); [Fair et al., 2008](#_ENREF_15); [Hagmann et al., 2008](#_ENREF_17)), voxels ([Eguiluz et al., 2005](#_ENREF_14); [van den Heuvel et al., 2008](#_ENREF_44); [Zuo et al., 2012](#_ENREF_48)) or electrode position ([Downes et al., 2012](#_ENREF_13)) as nodes, and activity associations such as cross correlation, mutual information and Granger causality as edges ([Bullmore and Sporns, 2009](#_ENREF_7); [Fair et al., 2008](#_ENREF_15); [Khazaee et al., 2015](#_ENREF_22); [Micheloyannis et al., 2009](#_ENREF_27); [Wang et al., 2010](#_ENREF_45)).

On the other hand, at the single cell level graphical models have been used to describe organizing principles of artificial neural networks, identifying neurons that could have a potential role orchestrating the overall network activity ([Iturria-Medina et al., 2008](#_ENREF_21); [Sporns, 2000](#_ENREF_38)). Such graphs are usually associated with a restricted set of parameters that describe the weight and direction of edges obtained by pairwise correlations, therefore are incapable of characterizing the emergent properties of cortical ensembles described by the whole population activity. Finally, although a few studies have applied graph theory to model network organization in calcium imaging data with single cell resolution in cultures or brain slices ([Bonifazi et al., 2009](#_ENREF_5); [Gururangan et al., 2014](#_ENREF_16); [Yatsenko et al., 2015](#_ENREF_46)), it has not been applied to define the functional connectivity based on the joint probability distribution of neuronal ensembles in awake behaving animals.

Cortical ensembles in primary visual cortex consist of strongly interconnected neurons ([Carrillo-Reid et al., 2016](#_ENREF_10); [Ko et al., 2011](#_ENREF_23)), forming a network structure that can be naturally modeled with graph theory, where nodes and edges are biologically meaningful, representing neurons and their connections respectively. We demonstrate that graph theory applied to CRFs allows the identification of cortical ensembles associated with different experimental conditions opening the possibility of targeting, with two-photon optogenetics, the most significant neurons from specific populations during physiological processes.

**Results**

**Identification of cortical ensembles from calcium imaging population data**

Cortical ensembles in primary visual cortex represent neuronal populations responding to specific features of visual stimuli ([Carrillo-Reid et al., 2015a](#_ENREF_8); [Carrillo-Reid et al., 2016](#_ENREF_10); [Miller et al., 2014](#_ENREF_28)). The overall activity of multiple cells at a given time window can be understood as a multidimensional array of population vectors where vectors pointing to a similar space can be considered as a group (Figure 1). We previously showed that population vectors defining a group (i.e. a cortical ensemble) can be extracted from multidimensional arrays performing singular value decomposition (SVD) ([Carrillo-Reid et al., 2015a](#_ENREF_8)). Even though SVD can identify cortical ensembles reliably, it lacks from a structured model that allows the systematic study of changes in functional connectivity during different experimental conditions.

**CRFs models predict external stimuli**

CRFs model the conditional distribution *p*(**y**|**x**) for observed population activity **x** over all nodes and network states **y** with an associated graphical structure. Therefore, CRFs have been successfully applied in diverse areas such as news and finance ([Peng et al., 2011](#_ENREF_31); [Tang et al.](#_ENREF_42)), bioinformatics ([Li et al., 2008](#_ENREF_25); [Liu et al., 2006](#_ENREF_26); [Sato and Sakakibara, 2005](#_ENREF_34)), computer vision ([He et al., 2004](#_ENREF_18); [Sminchisescu et al., 2006](#_ENREF_37)) and natural language processing ([Choi et al., 2005](#_ENREF_12); [Lafferty et al., 2001](#_ENREF_24)).

In order to construct a structured model from population activity with single cell resolution, we used CRFs representing neurons and their functional connections as nodes and edges in a graph (Figure 2A). To obtain the probability estimation to observe different network states based on observed population vectors, we assume that observed activity events from each neuron were generated by nodes in a graph structure, and that each node can have two values: ‘0’ corresponding to non-activity, and an ‘1’ corresponding to neuronal activity. In this way nodes interact with each other by connecting edges, which have four possible combinations ‘00’, ‘01’, ‘10’, and ‘11’, depending on the values of the two nodes on the edge. The two values associated with nodes and the four values associated with edges are characterized by a set of parameters called node potentials (ϕ0, ϕ1) and edge potentials (ϕ00, ϕ01, ϕ10, ϕ11) correspondingly (Figure 2A). These parameters reflect the likelihood of individual values on each node and edge. The strength of synchronization between two neurons can then be represented by a synchronization strength term . Using part of the observation data, we obtained the model parameters and performed cross-validation on the withheld data. We then eliminated weak edges by the synchronization strength term using a threshold generated from models constructed from shuffled data. Therefore, the likelihood of the neuronal population exhibiting a specific activation pattern can be described by the normalized product of the corresponding nodes and edge potentials (Figure 2A).

To integrate the information of external stimulus with the observed data, we added a hidden node for each presented stimulus, and set it to ‘1’ when the corresponding stimulus was on and ‘0’ when the stimulus was off (Figure 2B). We then trained CRFs using the real data with hidden nodes. In this way, the nodes that are directly connected to the hidden nodes depict different visual stimuli (Figure 2C).

CRFs model the conditional probability of network states given the observations. Therefore, by treating visual stimuli as nodes and comparing the output likelihood of observing each stimulus, we were able to predict visual stimuli from observed data. For example, given two visual stimuli (horizontal or vertical drifting gratings), the likelihood corresponding to observing each stimulus is defined by: and ,. Thus, the relative likelihood, , can be used to predict the presented stimuli (Figures 2D-2F). To evaluate the prediction performance, we calculated three standard measurements from the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN): accuracy, defined as (TP+TN)/(TP+TN+FP+FN); precision, defined as TP/(TP+FP); and recall, defined as TP/(TP+FN). Using these measurements we demonstrated that CRFs are able to define neuronal ensembles representing specific features of visual stimuli as the orientation of drifting-gratings (Figure 2G-I; Figure S1).

**CRFs capture population properties of neuronal ensembles**

To investigate if the functional connectivity of cortical ensembles identified by CRFs represent emergent network properties that cannot be understood by correlations, we compared CRF graphs against graphs constructed from pairwise correlation values (CC graph) between neurons, using observed spikes along with hidden nodes that represent visual stimuli (Figure 3A) ([Bonifazi et al., 2009](#_ENREF_5); [Downes et al., 2012](#_ENREF_13); [Sadovsky and MacLean, 2014](#_ENREF_33); [Zuo et al., 2012](#_ENREF_48)). CC and CRF graphs show significantly different density (Figure 3B) as well as distribution of node degrees, local clustering coefficients and centrality (Figures 4C-4E). Both also show nonrandom structures indicated by a heavy tail, compared with Erdős-Rényi random graphs, which preserve the number of nodes and edges (Figures 4C-4E). Thus CRF and CC graphs differ from each other (Figure 4), indicating that CRF graphs describe network properties that are not detected in CC graphs.

**Identification of core neurons from CRF ensembles using graph theory**

Brain networks show both structural and functional modularization, in both macro-scale and micro-scale ([Achard et al., 2006](#_ENREF_2); [Bonifazi et al., 2009](#_ENREF_5); [Hagmann et al., 2008](#_ENREF_17); [He et al., 2007](#_ENREF_19); [Shimono and Beggs, 2015](#_ENREF_36); [Sporns et al., 2007](#_ENREF_39); [Stetter et al., 2012](#_ENREF_40); [Zuo et al., 2012](#_ENREF_48)). Network modularization is often characterized by local structures with high inter-connectivity, where a group of neurons shows dense physical or functional connections. Such structures can be described by different concepts such as cliques, communities ([Palla et al., 2005](#_ENREF_30)), hubs and modules ([Bullmore and Sporns, 2009](#_ENREF_7)).

Most of the neurons used to predict visual stimuli interact with the hidden nodes through direct or one-step indirect connections (Figure ?). However, in order to design close loop optogenetic experiments targeting specific neurons it is necessary to identify core elements that can efficiently represent each neuronal ensemble. According to the Hammersley Clifford theorem, the probability distribution of CRFs can be factored as the product of clique potentials of maximal cliques:

where maximal cliques are complete subgraphs (fully interconnected subgraphs) that cannot be extended by adding more nodes (Figure 4A). Therefore, maximal cliques can be considered as functional units in a graph. We then examined the maximal cliques that contain at least one node that has a direct connection with the hidden nodes (Figure 4B). CRF graphs show significantly different number and size of maximal cliques than Erdős-Rényi random graphs (Figures 4C and 4D), indicating those structures cannot be considered as random. On the other hand, the properties of maximal cliques computed in CRF graphs represent more efficiently core neurons than CC graphs (Figure S2). Therefore, maximal cliques from CRFs allow the identification of core neurons that can be considered as the most significant elements for a specific given condition.

Core neurons from cortical ensembles identified with CRFs differed from descriptive methods ([Carrillo-Reid et al., 2015b](#_ENREF_9); [Carrillo-Reid et al., 2016](#_ENREF_10)) of ensemble identification using singular value decomposition (SVD) (Figure 5A). The total number of core neurons was not significantly different (Figure 5B) whereas only a small fraction of them were detected with both methods (Figure 5C), suggesting that few neurons can be representative of cortical ensembles. On the other hand, core neurons identified with CRFs are composed of a mixed population of cells that are highly tuned to a specific orientation (high OSI) and cells with low orientation selectivity (Figure S3) indicating that cortical ensembles are not purely orientation selective cells.

We then evaluated the identified core neurons by their performance of predicting external visual stimuli presented to the mice. To make predictions with ensemble activity, we calculated the cosine similarity between population vectors ([Carrillo-Reid et al., 2015a](#_ENREF_8)) observed in real data compared to population vectors defined by CRFs. Similarity coefficients between CRFs population vectors and real data reproduces presented stimulus (Figure 5D), and is specific to the stimulus activation periods. Prediction statistics show the CRFs defined ensembles have similar values of prediction accuracy, precision and recall than SVD method (Figures 5E-5G). Finally, CRFs outperform other graph theory measurements that could be used to define the most important members of a network (Figure S4).

**Core neurons from cortical ensembles are optimal for external stimuli prediction**

We next investigated whether core neurons from cortical ensembles identified by CRFs represent the optimal group of neurons for the prediction of presented visual stimuli. To do this, we randomly resized identified cortical ensembles adding or removing elements from the group (Figures 6A-6D) and examined the prediction performance. The similarity between population vectors of resampled ensembles has a maximum value when ensemble size is unchanged (Figure 6A). Furthermore, the core neurons from cortical ensembles achieve the best accuracy, precision and recall when predicting the presented visual stimuli, compared with resized ensembles (Figures 6B-6D).

So far we have shown that core neurons identified by CRFs represent the optimal population to predict external visual stimuli. This fact raises the question of whether CRF ensembles are a specific non-random subgroup. To answer this question, we randomly sampled a subset of the total neuronal population, ranging from 10% to 100% of the number of core neurons. We observed that prediction performance from random groups of neurons is significantly lower than CRF ensemble performance (Figures 6E-6H), indicating that identified ensembles are non-random groups of neurons.

**CRFs as a tool for the targeted manipulation of cortical microcircuits**

A challenging issue regarding the design of closed-loop optogenetic experiments to manipulate behavioral tasks in awake animals is the identification of neurons that could be used to recall learned patterns. It has been recently shown that the repetitive activation of an identified neuronal population with two-photon optogenetics imprints an artificial cortical ensemble that can be recalled later on by specific members of the ensemble ([Carrillo-Reid et al., 2016](#_ENREF_10)).

To investigate if core neurons identified by CRFs represent strongly connected neurons capable to recall a whole ensemble we used two-photon single cell stimulation to target nodes with high edge potential values (Figure 7A). The activation of highly connected neurons was able to induce pattern completion of imprinted ensembles (Figure 7B) whereas neurons with low connectivity were unable to recall imprinted ensembles (Figure 7C) demonstrating that CRFs could be used to target specific members of cortical ensembles that play a key role in the computational properties of cortical microcircuits.

**CRFs capture changes in network dynamics induced by circuit reconfiguration**

To investigate if our approach can describe changes in functional connectivity of cortical microcircuits after ensemble imprinting, we compared the models generated by CRFs before and after two-photon population manipulation of a given set of neurons (Figure 8A). Interestingly, the general properties of full models depicted during ongoing activity remain the same, suggesting that the imprinted ensemble has been added to the cortical microcircuit but preserving the overall network architecture (Figure 8B).

It has been suggested that the imprinting protocol reconfigures the connectivity between stimulated cells ([Carrillo-Reid et al., 2016](#_ENREF_10)). We used CRFs to measure the changes in the functional connectivity from the targeted group of neurons (Figure 8C). Our approach was able to describe specific changes in the connections that reflect the existence of the imprinted ensemble (Figure 8D). The fact that CRFs were able to describe changes in the reconfiguration of a specific subpopulation of neurons demonstrates the potential of structured prediction methods to study the fine modulation of neuronal microcircuits induced by pathological conditions.

**Discussion**

**Functional connectivity in cortical microcircuits**

In the past few decades, graph theory has been applied to characterize the structure and function of neuronal networks ([Achard and Bullmore, 2007](#_ENREF_1); [Bettencourt et al., 2007](#_ENREF_4); [Chiang et al., 2016](#_ENREF_11); [Downes et al., 2012](#_ENREF_13); [Fair et al., 2008](#_ENREF_15); [Hagmann et al., 2008](#_ENREF_17); [Iturria-Medina et al., 2008](#_ENREF_21); [Oh et al., 2014](#_ENREF_29); [Supekar et al., 2008](#_ENREF_41); [Yu et al., 2008](#_ENREF_47); [Zuo et al., 2012](#_ENREF_48)). While most of these studies operated on functional recording across multiple brain regions ([Achard and Bullmore, 2007](#_ENREF_1); [Chiang et al., 2016](#_ENREF_11); [Fair et al., 2008](#_ENREF_15); [Hinne et al., 2013](#_ENREF_20); [Zuo et al., 2012](#_ENREF_48)), only a few have focused on the general network properties of cortical circuits with recording from single neurons ([Bonifazi et al., 2009](#_ENREF_5); [Sadovsky and MacLean, 2014](#_ENREF_33); [Stetter et al., 2012](#_ENREF_40); [Yatsenko et al., 2015](#_ENREF_46)).

The majority of methods applied to infer the functional connectivity in brain slices ([Sadovsky and MacLean, 2014](#_ENREF_33); [Stetter et al., 2012](#_ENREF_40)) or *in vivo* ([Yatsenko et al., 2015](#_ENREF_46)) operate on the correlation matrix, and aim to recover the functional dependencies between observed neurons. Such methods are valuable for revealing network properties such as node degrees, clustering coefficients or functional hubs. However, these methods are model-free, therefore are incapable of describing the overall network dynamics based on the probability distribution of neuronal ensembles. In this study, we provide a tool for modeling the functional connectivity of mouse primary visual cortex *in vivo* using conditional random fields. Structured prediction methods not only reveal the structure of the functional connections, but also provide a full distribution of the conditional probability to find network states given the population activity, and are capable of predicting specific features of sensory stimuli (Figure 2).

**CRFs for identification of neuronal ensembles**

Compared with generative graphical models that capture the dependencies between all the possible variables from the model, CRFs only model sampled variables dependent on a given experimental condition. This is an advantage for classification tasks since discriminative models have better performance than generative models avoiding the exhaustive description of the joint probability distribution of observations as well as the assumptions of potentially complex dependencies between variables ([Lafferty et al., 2001](#_ENREF_24)). Additionally, given the finite number of network states described by population activity, the conditional distribution is sufficient for making predictions. Compared with other discriminative models such as Max entropy Markov model (MEMM), CRFs achieve higher accuracy by using global normalizers to overcome the local bias in MEMM induced by local normalizers ([Lafferty et al., 2001](#_ENREF_24)).

The difficulty of constructing CRFs lies in the computation of global normalizers. With an arbitrary graph structure, this problem is often intractable. Recent advances that combines Bethe free energy approximation and Frank-Wolfe methods for inference and learning model parameters allow fast and relatively accurate construction of cyclic CRFs ([Tang et al., 2016](#_ENREF_43)). Thus CRFs can be applied to datasets with hundreds of interconnected neurons. Constraints for applying CRFs still exist: with less than 400 samples of training data, the learned model does not achieve the best performance (Figure S5).

Compared with the previously used descriptive methods for neuronal ensemble identification ([Carrillo-Reid et al., 2015b](#_ENREF_9)), our approach modestly improved prediction accuracy. One reason could be that the current CRF learning algorithm executes separately the structure learning and parameter learning steps. Therefore, the learned graphical structure and parameters may not be the globally best matching ones. However, it is still computational unrealistic to explore all possible structures and parameter combinations. Additionally, approximations in parameter inference also does not guarantee global optimal.

**Physiological significance of targeted single cell optogenetic stimulation**

Electrical stimulation of visual cortex has been used for decades as an attempt to provide useful visual sensations to patients that have lost the functionality of their eyes ([Brindley and Lewin, 1968](#_ENREF_6)). The sensations produced by electrical stimulation of the visual cortex were termed phosphenes since they represented bright spots. A challenging issue regarding prostheses is the training of patients using devices with a large number of electrodes ([Shepherd et al., 2013](#_ENREF_35)). Our results suggest that after a given network have been trained the identification of core neurons could be used to recall learned patterns thus reducing the number of active points that require stimulation. The further development of network models based on population activity that can predict a given set of features embedded in visual stimuli will be crucial for the fine manipulation of cortical ensembles.

It has been shown that the connectivity of diverse systems described by graphs with complex topologies follow a scale-free power-law distribution ([Barabasi and Albert, 1999](#_ENREF_3)). Scale-free networks are characterized by the existence of a small subset of nodes with high connectivity ([Carrillo-Reid et al., 2015a](#_ENREF_8)). Similarly, cortical ensembles described by CRFs could be characterized by a subset of core neurons with strong synaptic connections. The existence of core neurons has been suggested in previous studies where perturbing the activity of single neurons was able to change the overall network dynamics ([Bonifazi et al., 2009](#_ENREF_5); [Carrillo-Reid et al., 2016](#_ENREF_10); [Hagmann et al., 2008](#_ENREF_17)).

To investigate the role of core neurons in a given cortical microcircuit during behavioral events, techniques for identifying and manipulating such neurons are needed. Our approach represents the first stage in the design of closed loop optogenetic experiments with single cell resolution.

**Figure legends**

**Figure 1. Cortical ensembles as a representation of multidimensional population vectors obtained with two-photon calcium imaging**

(A) Schematic representation of active neurons at different frames. Black dots represent active neurons at different times (left). Binary raster plot representing the overall network activity of observed neurons (right). Population vectors capture the coordinated activity of a given neuronal ensemble. (B) Population vectors can be understood as a multidimensional array in which clusters of population vectors taken from different times define network states..

**Figure 2. Classification of visual stimuli using CRFs**

(A) Schematic representation of CRFs. Shaded nodes (x) represent the observed binary spiking state of the neurons. White nodes (y) represent true states of the neurons, and are connected by edges that indicate their mutual dependencies; node potentials are defined over the two possible states of each node, and edge potentials are defined over the four possible states of each existing edge, depending on the state of the two nodes it connects. The probability distribution of the network over all possible states can therefore be factored using the node and edge potentials. (B) Illustration of hidden nodes (squares) in CRFs. In this case, two hidden nodes (squares) were added for the horizontal (red) and vertical (blue) visual stimuli, separately. Nodes that are directly connected to the two hidden nodes are also highlighted in the corresponding color. (C) Represetnative example of graphs constructed with hidden nodes. Nodes that are either directly connected to the hidden nodes or indirectly connected to the hidden nodes through one intermediate node are highlighted. Connecting edges between them are also highlighted. Square on the upper left corner corresponds to horizontal stimulus (red); square on the lower left corner corresponds to vertical stimulus (blue). (D) Example of relative log-likelihood calculated by CRFs. Black trace shows . Gray dashed line represents 0. Light red and blue vertical stripes indicate the time of horizontal and vertical stimuli, respectively. (E) Temporal course of ensemble classification. Top panel shows the prediction for horizontal stimulus; bottom panel shows the prediction for vertical stimulus. (F) Relative log-likelihood during horizontal and vertical stimuli. Red box represents the distribution of relative log-likelihood during horizontal stimulus; blue box represents the vertical stimulus. Gray region represents the threshold of 3 times baseline standard deviation level. (G-I) Accuracy (G), precision (H) and recall (I) of prediction for each orientation of visual stimuli.

**Figure 3. Graph properties of CRFs vs. correlation-based models**

(A) Representative examples of graphs constructed with CRFs and with pairwise correlations (CC). In the latter case, pairwise Pearson correlations between cells were calculated, and the threshold was determined by 5% significance level of correlation values of shuffled data. The node size is proportional to the node degree, and the edge color represents the synchrony edge potential (ф11+ ф00- ф01- ф10). (B) Graph density of CRFs and CC graphs. (C-E) Probability distribution of node degrees (C), local clustering coefficients (D), and eigenvector centrality (E) in CC and CRFs graphs. Both graphs differ from random models (black lines).

**Figure 4. Core neurons defined correlation-based models**

(A) Schematic representation of 3, 4 and 5-cliques with corresponding adjacency matrices. (B) Representative examples of maximal cliques that contain at least one node directly connected with hidden nodes. The top panel highlights all the maximal cliques for horizontal stimulus; the bottom panel highlights all the maximal cliques for vertical stimulus. (C) The number of maximal cliques from CRFs is significantly different from random graphs demonstrating that core neurons defined by maximal cliques are not a random group. (D) Probability distribution of maximal clique sizes from CRFs and random graphs.

**Figure 5. Predicting visual stimuli with identified core neurons**

(A) Examples of core neurons identified using SVD (green) and CRFs (orange). Circles represent core neurons of horizontal (left) and vertical (right) visual stimuli, respectively. Neurons shared between CRFs and SVD methods are represented by green and orange dots. (B) Percentage from the total population size representing core neurons for SVD and CRFs methods. (C) Percentage of shared neurons identified by SVD and CRFs methods. Percentage is calculated by number of cells that belong to both methods divided by the total number of unique cells in both methods. (D) Cosine similarity between population vectors that belong to given visual stimuli (color) compared to population vectors from different visual stimuli (black). (E) Accuracy, (F) precision and (G) recall of predictions for each visual stimuli using both methods.

**Figure 6. Core neurons efficiently represent presented visual stimuli**

(A) Cosine similarity between population vectors representing a given visual stimuli of randomly down-sampled or up-sampled core neurons (color). In gray is shown the cosine similarity of population vectors between different visual stimuli. (B) Accuracy, (C) precision and (D) recall of predictions from randomly down-sampled or up-sampled ensemble groups. (E) Cosine similarity between population vectors of randomly sampled cells from the whole population. (F) Accuracy, (G) precision and (H) recall of predictions from randomly chosen ensemble groups taken from the whole population.

**Figure 7. Highly connected core neurons have pattern completion properties**

(A) CRFs graph highlighting core neurons with high edge potential values (top). Two-photon optogenetic targeting of highly connected core neurons are able to recall imprinted ensembles (bottom). (B) Neurons with low edge potential values (top) were not able to induce pattern completion of imprinted ensembles (bottom).

**Figure 8. Network reconfiguration in a targeted population of neurons**

(A) CRFs graphs of ongoing activity before (left) and after (right) population photostimulation. (B) Measurements of the network parameters showing that the functional connectivity of the whole model remains stable after the imprinting protocol. (C) Edge potentials from optogenetically activated cells extracted from the full model before (left) and after (right) population photostimulation. (D) Changes in functional connectivity between targeted neurons captured by CRFs.

**Experimental Procedures**

***Conditional Random Fields***

We construct a conditional random field (CRF) with the observed population activity where , and the target hidden network state , where , for samples (time points). For each sample, the conditional probability can be expressed as:

where is a vector of sufficient statistics of the distribution, is a vector of parameters, and is the partition function:

The conditional probability can be factored over a graph structure , where is the collection of nodes representing observation variables and target variables, and is the collection of subsets of . The conditional dependencies can be then written as

This model is a generalized version of Ising models, which have been previously applied to model neuronal networks ([Yu et al., 2008](#_ENREF_47)). The log-likelihood of each observation can be then written as:

Given the inferred binary spikes from raw imaging data, we construct a CRF model by two steps: (1) structure learning, and (2) parameter learning. For structure learning, we learnt a graph structure using l1-regularized logistic regression and performed structure elimination by thresholding the edge potentials ([Ravikumar et al., 2010](#_ENREF_32)). Based on the learnt structure, we use the Bethe approximation to approximate the partition function, and iterative Frank-Wolfe methods to perform parameter estimation by maximizing the log-likelihood of the observations with a quadratic regularizer ([Tang et al., 2016](#_ENREF_43)):

Cross-validation was done to find the best model parameters via model likelihood.

***Shuffling Methods***

We performed shuffling of binary data while preserving the activity level for each cell and each frame. To do this, we randomly selected two cells and two time points where they show different activity (‘0’ and ‘1’), and exchange the activity pairs across them. This procedure was repeated for a large number (2n, where n is the number of total spikes) to complete one shuffling. The activity of the stimulus-representing hidden nodes were not shuffled. For each dataset, shuffling was done for 100 times unless noted otherwise.

***Edge Elimination of CRF Graphs***

Edges in the constructed CRF models can represent both positive correlation and negative correlation. To eliminate edges that do not represent strong synchronization between neurons, we trained CRF models with 100 shuffled data, and obtained the synchronization term . We then fitted the distribution of these synchronization terms to a normal distribution, and determined a threshold by bottom 30% quantile. All edges in the real-data CRF model that show a synchronization term below the threshold were eliminated.

***Correlation-based Graphs***

To construct correlation-based graphs, we first calculated the Pearson correlation coefficients between the binary spike vectors of each pair of cells. Then, we generated correlation threshold by calculating Pearson correlation coefficients of shuffled data, fitting the coefficients to a normal distribution, and finding the 95% CDF level.

***Maximal Cliques***

Finding maximal cliques using exhaustive search methods is computationally unrealistic with a relatively large number of vertices. To find maximal cliques in an adjacency matrix efficiently, we used the Bron-Kerbosch algorithm. This algorithm recursively detects all the maximal cliques in a given graph . The algorithm starts with three sets: an empty set with currently growing maximal clique, a set with all prospective vertices connected to all vertices in , and a set with nodes that have been processed. In each call of the algorithm, a pivot vertex with the largest node degree is chosen. Since for each vertex , either the vertex or its non neighbors but not both will be in a clique, each is tested as candidate component for by recursively calling the algorithm with and restricted to the neighbors of . The algorithm then moves from to , and reports as a maximal clique when both and are empty.

***Graph properties***

Given the adjacency matrix where if node is linked to node , we investigated the following graph properties: graph density, node degrees, local clustering coefficients, and eigenvector centrality. Graph density is calculated as the number of existing edges divided by the number of total possible edges:

where *NV* is the number of vertices in the graph. Node degree is defined for node as the number of edges connected to it:

.

Local clustering coefficient is defined for each node as the fraction edges connected to it over the total number of possible edges between the node's neighbors (nodes that have a direct connection with it). Eigenvector centrality is defined on the relative centrality score matrix , where

This can be written in the form of eigenvector equation:

Solving the above equation gives a set of eigenvalues and associated eigenvectors. The entry of the eigenvector associated with the largest gives the eigenvector centrality for node .

***Prediction with cosine similarity***

Identified core neurons were represented by a binary vector over all neurons, and the entries corresponding to the ensemble members were set to 1, while the rest were set to 0. Cosine similarities between ensemble vectors and frame activity vectors were calculated, and a threshold was determined by 3 times the standard deviation of baseline noise. The cosine similarity between two frame activity vectors depicts the angle between two vectors in the high-dimensional space; orthogonal angles indicate that the active neurons in the two frames are mostly different, while small angles indicate that the active neurons are mostly the same. Frames that were significantly similar compared with the threshold were taken as stimulus-on frames.

**References**

Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. PLoS Comput Biol *3*, e17.

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci *26*, 63-72.

Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. Science *286*, 509-512.

Bettencourt, L.M., Stephens, G.J., Ham, M.I., and Gross, G.W. (2007). Functional structure of cortical neuronal networks grown in vitro. Phys Rev E Stat Nonlin Soft Matter Phys *75*, 021915.

Bonifazi, P., Goldin, M., Picardo, M.A., Jorquera, I., Cattani, A., Bianconi, G., Represa, A., Ben-Ari, Y., and Cossart, R. (2009). GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. Science *326*, 1419-1424.

Brindley, G.S., and Lewin, W.S. (1968). The sensations produced by electrical stimulation of the visual cortex. J Physiol *196*, 479-493.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews Neuroscience *10*, 186-198.

Carrillo-Reid, L., Lopez-Huerta, V.G., Garcia-Munoz, M., Theiss, S., and Arbuthnott, G.W. (2015a). Cell Assembly Signatures Defined by Short-Term Synaptic Plasticity in Cortical Networks. Int J Neural Syst *25*, 1550026.

Carrillo-Reid, L., Miller, J.-E.K., Hamm, J.P., Jackson, J., and Yuste, R. (2015b). Endogenous sequential cortical activity evoked by visual stimuli. The Journal of neuroscience : the official journal of the Society for Neuroscience *35*, 8813-8828.

Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D.S., and Yuste, R. (2016). Imprinting and recalling cortical ensembles. Science *353*, 691-694.

Chiang, S., Cassese, A., Guindani, M., Vannucci, M., Yeh, H.J., Haneef, Z., and Stern, J.M. (2016). Time-dependence of graph theory metrics in functional connectivity analysis. Neuroimage *125*, 601-615.

Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing HLT 05, 355-362.

Downes, J.H., Hammond, M.W., Xydas, D., Spencer, M.C., Becerra, V.M., Warwick, K., Whalley, B.J., and Nasuto, S.J. (2012). Emergence of a small-world functional network in cultured neurons. PLoS Comput Biol *8*, e1002522.

Eguiluz, V.M., Chialvo, D.R., Cecchi, G.A., Baliki, M., and Apkarian, A.V. (2005). Scale-free brain functional networks. Phys Rev Lett *94*, 018102.

Fair, D.A., Cohen, A.L., Dosenbach, N.U., Church, J.A., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., and Schlaggar, B.L. (2008). The maturing architecture of the brain's default network. Proc Natl Acad Sci U S A *105*, 4028-4032.

Gururangan, S.S., Sadovsky, A.J., and MacLean, J.N. (2014). Analysis of graph invariants in functional neocortical circuitry reveals generalized features common to three areas of sensory cortex. PLoS Comput Biol *10*, e1003710.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. PLoS Biol *6*, e159.

He, X., Zemel, R.S., and Carreira-Perpinan, M.A. (2004). Multiscale conditional random fields for image labeling. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition *2*, 695 -702.

He, Y., Chen, Z.J., and Evans, A.C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. Cereb Cortex *17*, 2407-2419.

Hinne, M., Heskes, T., Beckmann, C.F., and van Gerven, M.A.J. (2013). Bayesian inference of structural brain networks. NeuroImage *66*, 543-552.

Iturria-Medina, Y., Sotero, R.C., Canales-Rodriguez, E.J., Aleman-Gomez, Y., and Melie-Garcia, L. (2008). Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. Neuroimage *40*, 1064-1076.

Khazaee, A., Ebrahimzadeh, A., and Babajani-Feremi, A. (2015). Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology *126*, 2132-2141.

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

Lafferty, J., McCallum, A., and Pereira, F.C.N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning *8*, 282-289.

Li, C.T., Yuan, Y., and Wilson, R. (2008). An unsupervised conditional random fields approach for clustering gene expression time series. Bioinformatics *24*, 2467-2473.

Liu, Y., Carbonell, J., Weigele, P., and Gopalakrishnan, V. (2006). Protein fold recognition using segmentation conditional random fields (SCRFs). J Comput Biol *13*, 394-406.

Micheloyannis, S., Vourkas, M., Tsirka, V., Karakonstantaki, E., Kanatsouli, K., and Stam, C.J. (2009). The influence of ageing on complex brain networks: a graph theoretical analysis. Hum Brain Mapp *30*, 200-208.

Miller, J.E., Ayzenshtat, I., Carrillo-Reid, L., and Yuste, R. (2014). Visual stimuli recruit intrinsically generated cortical ensembles. Proc Natl Acad Sci U S A *111*, E4053-4061.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M.*, et al.* (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207-214.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature *435*, 814-818.

Peng, H.-K., Zhu, J., Piao, D., Yan, R., and Zhang, Y. (2011). Retweet Modeling Using Conditional Random Fields. In 2011 IEEE 11th International Conference on Data Mining Workshops (IEEE), pp. 336-343.

Ravikumar, P., Wainwright, M.J., and Lafferty, J.D. (2010). High-dimensional Ising model selection using ℓ1-regularized logistic regression. The Annals of Statistics *38*, 1287-1319.

Sadovsky, A.J., and MacLean, J.N. (2014). Mouse visual neocortex supports multiple stereotyped patterns of microcircuit activity. J Neurosci *34*, 7769-7777.

Sato, K., and Sakakibara, Y. (2005). RNA secondary structural alignment with conditional random fields. Bioinformatics *21 Suppl 2*, ii237-242.

Shepherd, R.K., Shivdasani, M.N., Nayagam, D.A., Williams, C.E., and Blamey, P.J. (2013). Visual prostheses for the blind. Trends in biotechnology *31*, 562-571.

Shimono, M., and Beggs, J.M. (2015). Functional Clusters, Hubs, and Communities in the Cortical Microconnectome. Cereb Cortex *25*, 3743-3757.

Sminchisescu, C., Kanaujia, A., and Metaxas, D. (2006). Conditional models for contextual human motion recognition. Computer Vision and Image Understanding *104*, 210-220.

Sporns, O. (2000). Theoretical Neuroanatomy: Relating Anatomical and Functional Connectivity in Graphs and Cortical Connection Matrices. Cerebral Cortex *10*, 127-141.

Sporns, O., Honey, C.J., and Kotter, R. (2007). Identification and classification of hubs in brain networks. PLoS One *2*, e1049.

Stetter, O., Battaglia, D., Soriano, J., and Geisel, T. (2012). Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. PLoS Comput Biol *8*, e1002653.

Supekar, K., Menon, V., Rubin, D., Musen, M., and Greicius, M.D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput Biol *4*, e1000100.

Tang, K., Gubert, H., Tonge, R., Wang, A., Wu, L., Campbell, D., Kedzie, C., Wang, L., Russell, A., and Kimball, A. Learning a Graphical Model of Bloomberg Financial and News Data.

Tang, K., Ruozzi, N., Belanger, D., and Jebara, T. (2016). Bethe Learning of Graphical Models via MAP Decoding. Artificial Intelligence and Statistics (AISTATS).

van den Heuvel, M.P., Stam, C.J., Boersma, M., and Hulshoff Pol, H.E. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. Neuroimage *43*, 528-539.

Wang, J., Zuo, X., and He, Y. (2010). Graph-based network analysis of resting-state functional MRI. Front Syst Neurosci *4*, 16.

Yatsenko, D., Josic, K., Ecker, A.S., Froudarakis, E., Cotton, R.J., and Tolias, A.S. (2015). Improved estimation and interpretation of correlations in neural circuits. PLoS Comput Biol *11*, e1004083.

Yu, S., Huang, D., Singer, W., and Nikolic, D. (2008). A small world of neuronal synchrony. Cereb Cortex *18*, 2891-2901.

Zuo, X.N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F.X., Sporns, O., and Milham, M.P. (2012). Network centrality in the human functional connectome. Cereb Cortex *22*, 1862-1875.