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

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

Neurotechnology Center, Department of Biological Sciences and Neuroscience, Columbia University, New York, NY, 10027.

**Corresponding author:**

Rafael Yuste, M.D., Ph.D. [rafaelyuste@columbia.edu](mailto:rafaelyuste@columbia.edu)

906 NWC Building 550 West 120 Street, Box 4822, New York, NY, 10027.

**Author Contributions:** L.C.-R & S.H. contributed equally to this work. Conceptualization, L.C.-R; Methodology, L.C.-R.,S.H.,T.J.; Investigation, L.C.-R.; Writing – Original Draft, L.C.-R. & S.H.; Writing – Review & Editing, L.C.-R., S.H., T.J., & R.Y.; Resources, L.C.-R., S.H. & T.J.; Funding Acquisition, T.J. & R.Y.

**Acknowledgments:** Laboratory members for valuable comments. Laboratory technicians for virus injections. Stanford Neuroscience Gene Vector and Virus Core for AAVdj virus. Defense Advanced Research Projects Agency-Department of Defense (SIMPLEX N66001-15-C-4032). S.H. is a Howard Hughes Medical Institute International Student Research fellow. The authors declare no competing financial interests.

**Highlights <85**

- CRFs infer functional connectivity of cortical ensembles for different visual stimuli **85**

- CRF model allows identification of optimal cortical ensembles **61**

- High-ranked neurons from CRF models have pattern completion capability **67**

- CRFs capture network reconfiguration induced by two-photon optogenetic stimulation **82**

**Summary 143<150**

The prediction of different states from many variables with mutual dependencies is fundamental for the creation of models in varied applications including natural language, computer vision and bioinformatics. Such problem 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, CRF’s application to infer the functional connectivity of biological neural networks remains unexplored. We used CRFs in population calcium imaging from primary visual cortex (V1) of awake head-fixed mice to predict visual stimuli and identify cortical ensembles. Using simultaneous two-photon imaging and optogenetics we show that our approach can be used to identify high-ranked neurons from artificially imprinted ensembles with pattern completion capability. Our method provides a resource to study functional changes in overall network activity evoked by single cell targeting of cortical ensembles.

**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 animals ([Carrillo-Reid et al., 2016](#_ENREF_10)). However, how the activation of specific groups of neurons relate to the function of cortical microcircuits has been difficult to elucidate because it requires the identification of single cells 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 optimal network structure underlying 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 optimal configuration of neuronal ensembles that allows the prediction of different visual stimuli in awake 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 CRFs allow the identification of cortical ensembles associated with different experimental conditions, opening the possibility of targeting with single cell resolution the most significant neurons from specific populations during physiological processes.

**Results**

**Functional connectivity inference from calcium imaging population data using CRFs**

CRFs model the conditional distribution *p*(**y**|**x**), where **x** represents observations and **y** represents true labels associated with a graphical structure. Since no assumptions are made on **x**, CRFs can accurately describe the conditional distribution with complex dependencies in observation variables. Therefore, CRFs have been successfully applied in diverse areas such as news ([Peng et al., 2011](#_ENREF_31)), 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 1A). To obtain the probability estimation to observe different network states based on observed population vectors (Figure S1), 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 and edge potentials correspondingly (Figure 1A). These parameters reflect the likelihood of individual values on each node and edge. Using part of the observation data, we obtained the model parameters and performed cross-validation on the withheld data. Therefore, the normalized product of the corresponding nodes and edge potentials describes the likelihood of the neuronal population that has a specific activation pattern.

To integrate the information of external stimulus with the observed data, we added a variable for each presented stimulus, and set it to ‘1’ when the corresponding stimulus was on and ‘0’ when the stimulus was off (Figures 1A and 1B). The general properties of CRF models obtained with added variables do not significantly differ from CRF models obtained without added variables (Figure S2). CRFs model the conditional probability of network states given the observations. Therefore, by treating visual stimuli as added variables and comparing the output likelihood of observing each stimulus, we were able to predict visual stimuli from observed data. In this way, the nodes that are directly connected to the added variables represent different visual stimuli. 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 1C and 1D). 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 infer specific features of visual stimuli as the orientation of drifting-gratings (Figures 1E-H; Figure S3). Moreover, such CRFs constructed with a minimum of 400 frames given a population of 100 neurons show stable prediction performance (Figure S4).

**Identification of optimal cortical ensembles from CRFs**

Cortical ensembles in primary visual cortex represent neuronal populations with modular properties ([Carrillo-Reid et al., 2015a](#_ENREF_8); [Carrillo-Reid et al., 2016](#_ENREF_10); [Miller et al., 2014](#_ENREF_28)). Neural networks show both structural and functional modularization, in both macro-scale and micro-scale levels ([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)).

In order to design close loop optogenetic experiments with single cell resolution, it is necessary to identify optimal cortical ensembles that can efficiently represent different visual stimuli. The classifier nature of CRFs provides a convenient way of defining optimal cortical ensembles. For each neuron, we set its activity to be either ‘1’ or ‘0’ in all population activity vectors of the dataset, and compared the output likelihood with the trained CRF models (Figure 2A). Then, we calculated single neuron preference by binarizing the likelihood difference (Figure 2B) and counting the number of predictions under each visual stimulus (Figure 2C). In this way we defined the optimal cortical ensembles as the neuronal ensembles that prefer each visual stimulus (Figure 2C and 2D). Additionally, our method can be robustly extended to datasets with more visual stimuli types and different experimental settings (e.g. Allen Brain Observatory datasets) (Figure S5).

We next investigated whether 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 3A-3D) and examined the prediction performance. The similarity between population vectors of resampled ensembles has a maximum value when ensemble size is unchanged (Figure 3A). Furthermore, the neurons from optimal cortical ensembles achieve the best accuracy, precision and recall when predicting the presented visual stimuli, compared with resized ensembles (Figures 3B-3D).

So far we have shown that 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 optimal neurons. We observed that prediction performance from random groups of neurons is significantly lower than CRF ensemble performance (Figures 3E-3H), indicating that optimal ensembles are non-random groups of neurons.

Optimal neurons from cortical ensembles identified with CRFs have similar properties and prediction performance to previously used dimensional reduction methods ([Carrillo-Reid et al., 2015b](#_ENREF_9); [Carrillo-Reid et al., 2016](#_ENREF_10)) for ensemble identification (Figure S3), suggesting that few neurons can be representative of cortical ensembles. On the other hand, optimal neurons identified with CRFs are composed of a mixed population of cells with high orientation selectivity index (OSI) and cells with low orientation selectivity (Figure S3) indicating that cortical ensembles are not purely orientation selective cells. Therefore, CRFs allow the identification of optimal cortical ensembles that can be considered as the most significant elements for a specific experimental condition.

**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 cortical ensembles imprinted in neuronal microcircuits. 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 CRFs are able to identify neurons with pattern completion capability, we computed the node strength as the summation of edge potentials in all connecting edges for each node from the CRF models, and selected high-ranked neurons with strong node strength. We then used two-photon single cell stimulation to target high-ranked neurons (Figure 4A). Single cell optogenetic stimulation of high-ranked neurons (Figures 4B and 4C) was able to evoke pattern completion of imprinted ensembles (Figure 4D) whereas low-ranked neurons were unable to recall imprinted ensembles (Figure 4E), demonstrating that CRFs could be used to target single neurons that play a key role in the computational properties of cortical microcircuits.

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

To investigate if our approach can describe changes in functional connectivity of cortical microcircuits evoked by network perturbations, we compared the models generated by CRFs before and after two-photon population manipulation of a given set of neurons for several times, a maneuver that build new networks in the brain ([Carrillo-Reid et al., 2016](#_ENREF_10)). Our results demonstrated the reconfiguration of the functional connectivity between cortical ensembles after a new ensemble has been imprinted in the cortex (Figure 5A) Interestingly, the general properties of CRFs models before and after population photostimulation remain stable, suggesting that the imprinted ensemble has been added to the cortical microcircuit but preserving a balance with the overall network structure (Figures 5B-5F).

The fact that CRFs were able to describe changes in the reconfiguration of cortical neurons demonstrates the potential of structured prediction methods to study the modulation of neuronal microcircuits induced by external perturbations or 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 recordings 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 1).

**CRFs for identification of neuronal ensembles**

Compared with generative models that make assumptions on the dependencies between all the observation variables from the model, CRFs only model the hidden system states dependent on observed features, which can be either binary or continuous. No independence assumptions are made between observed variables, therefore CRFs avoid potential errors under these assumptions introduced by unobserved common inputs of the neuronal population. Additionally, given the finite number of network states described by population activity, the conditional distribution is sufficient for making predictions, both for the population state and for identifying representative cells in each state. One popular example of generative models for functional connectivity is the Ising model or the more generalized Potts model ([Tavoni et al., 2016](#_ENREF_43); [Yu et al., 2008](#_ENREF_47)). The generative nature of these models renders their ability to integrate complex dependencies between variables. Compared with other discriminative finite-state models such as Maximum Entropy Markov Models (MEMM), CRFs use global normalizers to overcome the local bias in MEMM induced by local normalizers, and have been shown to achieve higher accuracy in applications such as sequence labeling ([Lafferty et al., 2001](#_ENREF_24)). Therefore, CRFs are ideal for modeling cortical functional connectivity and for identifying optimal ensembles.

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_42)). 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 S4).

Compared with the previously used descriptive methods for neuronal ensemble identification ([Carrillo-Reid et al., 2015b](#_ENREF_9)), our approach modestly improved prediction accuracy (Figure S3). 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.

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 S1). 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.

**Physiological significance of single cell targeting with 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 neurons with pattern completion capability could be used to reduce 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 neurons with strong synaptic connections. The existence of neurons with pattern completion capability 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)).

Our approach demonstrated the importance of single cell optogenetic manipulation in the design of closed loop experiments to investigate the role of a specific subpopulation of neurons in a given cortical microcircuit during different behavioral events.

**Figure legends**

**Figure 1. Classification of visual stimuli using CRFs**

(A) Schematic representation of CRFs. Circles represent neurons. Squares represent added nodes depicting visual stimuli. 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 factorized using the node and edge potentials. (B) Example of a CRF graph constructed from real data. In this case, added nodes (squares) represent horizontal (red) and vertical (blue) drifting-gratings. First- and second-degree connections with the added nodes are highlighted in the corresponding color. Scale bar represents 50μm. (C) Relative likelihood calculated by CRFs. Gray region represents 3 S.D. Colored stripes indicate visual stimuli. Scale bar represents 200 frames. (D) Temporal course of ensemble classification for horizontal (top) and vertical (bottom) drifting-gratings. Scale bar represents 10 seconds. (E) Mean likelihood difference during horizontal and vertical visual stimulus. (F) Accuracy, (G) precision and (H) recall for vertical and horizontal drifting-gratings.

**Figure 2. Identification of cortical ensembles with CRFs**

(A) Schematic representation of ensemble identification from CRF models. The activity of the ith neuron is set to ‘1’ or ‘0’ at each frame, and the likelihood pmi,1 and pmi,0 of modified population vectors is calculated. (B) An example of likelihood difference (bottom panel) and predictions (top panel) for high activity frames from a neuron belonging to a cortical ensemble representing horizontal drifting-gratings (red). (C) Prediction ratio of neurons belonging to optimal cortical ensembles for horizontal (red) and vertical (blue) drifting-gratings are represented by colors. (D) Spatial map of optimal cortical ensembles from CRFs for horizontal (red) and vertical (blue) drifting-gratings. Scale bar represents 50μm.

**Figure 3. CRF cortical ensembles are optimal for visual stimuli identification**

(A) Cosine similarity between population vectors representing optimal cortical ensembles identified with CRFs for a given visual stimuli. Each optimal cortical ensemble was randomly down-sampled or up-sampled (orange). The cosine similarity of population vectors belonging to different visual stimuli is shown in gray. (B) Accuracy, (C) precision and (D) recall of predictions from randomly down-sampled or up-sampled optimal cortical ensembles. (E) Cosine similarity between population vectors of randomly sampled cells. (F) Accuracy, (G) precision and (H) recall of predictions from randomly chosen ensembles.

**Figure 4. High-ranked neurons from CRF models have pattern completion capability**

(A) Schematic representation of high-ranked neurons from CRF models. High-ranked neurons are nodes where the summation of edge potential values is maximal. (B) CRF graphical model from simultaneous two-photon imaging and two-photon optogenetic single cell stimulation. Squares depict added nodes representing 10 stimulation trials of different neurons. High-ranked neurons are highlighted in red. Edge potential color tone represents the strength of ; node color represents the rank of cells. Scale bar represents 50μm. (C) Rank distribution of the whole population. (D) Two-photon optogenetic single cell stimulation of targeted high-ranked neurons was able to recall imprinted ensembles demonstrating pattern completion capability. (E) Single cell optogenetic stimulation of low-ranked neurons was not able to induce pattern completion of imprinted ensembles. Scale bar represents 50μm.

**Figure 5. Reconfiguration of cortical microcircuits induced by two-photon optogenetic population manipulation**

(A) CRF graphical models from simultaneous two-photon imaging and two-photon optogenetic single cell stimulation of a high-ranked neuron with pattern completion capability before (left) and after (right) two-photon optogenetic population imprinting. Square on bottom left represents added node for single cell stimulation of 10 trials. Edge color tone represents the strength of ; node color represents the rank of cells. Node size represents the node degree. Scale bar represents 50μm. (B) Graph density, (C) node strength, (D) node degree, (E) clustering coefficient and (F) centrality values before and after imprinting protocol remained stable indicating that the imprinted ensemble reconfigured the network structure preserving the global network properties (n = 74 neurons; paired *t*-test; n.s. p>0.05).

**Experimental Procedures**

***Datasets***

Unless otherwise indicated, datasets in this study is previously published by the same group ([Carrillo-Reid et al., 2016](#_ENREF_10)). This dataset contains two-photon calcium imaging from GCaMP6s-expressing mice through AVV virus, and two-photon photostimulation of C1V2/GCaMP6s co-expressing mice in primary visual cortex. In the former dataset, mice were shown with horizontal and vertical drifting gratings. Spikes were extracted by binarizing dF/F traces with 3 standard deviations (S.D.). To reduce the bias caused by time points where the population show no or minimum firing rates, only high-activity frames with spike number above shuffled data were kept for further analysis.

We also used a publicly available dataset from the Allen Brain Observatory (<http://observatory.brain-map.org/visualcoding>) along with the SDK for extracting related information (<http://alleninstitute.github.io/AllenSDK/>) by Allen Institute of Brain Science. We selected 6 experiments with Cux2-CreERT2 driving GCaMP6f expression, imaging area VISp (primary visual cortex), imaging depth 175μm, and drifting gratings stimuli with 8 orientations and directions; the corresponding experiment container IDs are 511507650, 511509529, 511510650, 511510670, 511510718, and 511510855. To extract spikes, we calculated the dF/F from neuropil corrected traces, and binarized with 3 S.D. High-activity frames were kept as above. Drifting gratings with the same orientation but opposite directions were treated as the same.

***Conditional Random Fields***

We construct a conditional random field (CRF) with the feature vector where , and the target binary population activity vector , 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_42)):

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

***Identifying optimal cortical ensembles***

To find optimal cortical ensembles for each condition, we iterate through all the neurons and identify their contribution in predicting stimulus conditions with the population. To this end, for the neuron in population, we set its activity to be ‘1’ and ‘0’ in turn, in all M frames. With the two resulting population vectors in the frame among all samples, we calculate the likelihood of them coming from the trained CRF model:

Then, we take the likelihood difference vector

and threshold it by baseline plus 3 times the standard deviation of noise. This gives a binary prediction vector for each neuron over the entire experiment. We then counted the prediction ratio in each condition:

where denotes the number of predictions during stimulus for neuron , and denotes the number of frames in stimulus . Cortical ensemble for each visual stimulus is defined as the set of neurons that shows the highest prediction ratio in the corresponding visual stimulus. To exclude neurons with low predictive ability, we threshold with a value obtained from cross-validation, using the accuracy of cosine similarity prediction as readout.

***Prediction with cosine similarity***

Identified optimal cortical ensembles 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 baseline plus 3 times the standard deviation of 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.

***High-ranked cells***

We define the node strength as the sum of the ‘11’ term of edge potentials from all connecting edges:

Here denotes the number of connecting edges for node . The defined rank reflects the importance of a given cell in co-activating with other cells.

***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 .

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

Tavoni, G.A.-O.h.o.o., Cocco, S., and Monasson, R. (2016). Neural assemblies revealed by inferred connectivity-based models of prefrontal cortex recordings. Journal of computational neuroscience.

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.