**Identification and Targeting of Cortical Ensembles with Graphical Models**

Luis Carrillo-Reid1,2, Shuting Han1,2, Tony Jebara3 and Rafael Yuste1,2.

1 Departments of Biological Sciences, Columbia University, New York, NY 10027, USA

2 Neurotechnology Center, Columbia University, New York, NY 10027, USA

3 Department of Computer Science, Columbia University, New York, NY, 10027, USA

**Corresponding authors:**

Luis Carrillo-Reid. Ph.D. lc2998@columbia.edu

Rafael Yuste, M.D., Ph.D. 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:** We thank Ekaterina Taralova for pioneering the project, laboratory members for valuable comments and virus injections, Columbia Yeti shared High Performance Computing Cluster for computation resources, Stanford Neuroscience Gene Vector and Virus Core for AAVdj virus and the NEI (DP1EY024503, R01EY011787) and DARPA SIMPLEX N66001-15-C-4032 for funding. S.H. is a Howard Hughes Medical Institute International Student Research fellow. The authors declare no competing financial interests.

**Highlights**

- Graphical Conditional Random Field (CRFs) models identify cortical ensembles

- CRFs infer functional connectivity for different visual stimuli

- High-ranked neurons in CRF models have pattern completion capability

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

**Summary**

A fundamental problem for machine learning in applications including natural language, computer vision and bioinformatics is the prediction of internal states from measuring variables with mutual dependencies. One solution is structured prediction methods that combine graphical models and classification algorithms, such as Conditional random fields (CRFs) models. However, the application of CRF’s to infer the functional connectivity of biological neural networks remains relatively unexplored. We used CRFs in population calcium imaging data from primary visual cortex (V1) of awake head-fixed mice to identify cortical ensembles and predict visual stimuli. Using simultaneous two-photon imaging and optogenetics we identify high-ranked neurons from artificially imprinted ensembles with pattern completion capability. Our method can also capture functional changes in network activity evoked by single cell targeting. CRFs models could be broadly used for deciphering the structure and function of neural circuits.

**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 (Miller et al., 2014). Recent advances in two-photon calcium imaging and two-photon optogenetics have made possible the imprinting and recalling of cortical ensembles with single cell resolution in awake animals (Carrillo-Reid et al., 2016). However, how the activation of specific groups of neurons relate to the function of cortical microcircuits has been difficult to elucidate. This is partly because it requires the online identification of single cells that can be targeted during close-loop optogenetic experiments, potentially under interventional manipulation of learned behavioral tasks.

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

In addition, 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; Sporns, 2000). 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, 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; Gururangan et al., 2014; Yatsenko et al., 2015), but these methods have 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; Ko et al., 2011), 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. Here, 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**) of a network, 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 associated with a graphical structure that is used to constrain the interdependencies between labels. Therefore, CRFs have been successfully applied in diverse areas of machine learning such as news (Peng et al., 2011), bioinformatics (Li et al., 2008; Liu et al., 2006; Sato and Sakakibara, 2005), computer vision (He et al., 2004; Sminchisescu et al., 2006) and natural language processing (Choi et al., 2005; Lafferty et al., 2001).

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, respectively (Figure 1A). To estimate the probability of different network states from 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 are also known as potential functions and reflect the scores of individual values on each node and edge. Using part of the observation data, we first estimate model parameters and then perform cross-validation on held-out data. The final normalized product of the corresponding nodes and edge potentials describes the likelihood that the neuronal population exhibits a specific activation pattern.

To integrate information of the external stimulus along with the observed neuronal data, we added an additional node for each type of stimulus that was presented to the animal. This node was set to ‘1’ when the corresponding stimulus was on and ‘0’ when the stimulus was off (Figures 1A and 1B). The general and mathematical properties of CRF models obtained with added nodes did not significantly differ from CRF models obtained without added nodes and CRFs modeled the conditional probability of network states given the observations (Figure S2). Therefore, by treating visual stimuli as added nodes and comparing the output likelihood of observing each stimulus, we were able to predict visual stimuli from observed data. In this way, the nodes directly connected to the added nodes 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 classify the presented stimuli (Figures 1C and 1D). To evaluate the classification performance, we examined the standard receiver operating characteristic (ROC) curve of the two stimuli (Figure 1E) as well as the area under curve (AUC, Figure 1G). We also 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-1J).

**Identification of cortical ensembles using CRFs and stimulus prediction**

Neural networks show both structural and functional modularization, in both macro-scale and micro-scale levels (Achard et al., 2006; Bonifazi et al., 2009; Hagmann et al., 2008; He et al., 2007; Shimono and Beggs, 2015; Sporns et al., 2007; Stetter et al., 2012; Zuo et al., 2012). 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), hubs and modules (Bullmore and Sporns, 2009).

In primary visual cortex, coactive cortical ensembles represent neuronal populations with modular properties (Carrillo-Reid et al., 2015a; Carrillo-Reid et al., 2016; Cossart et al., 2003; Mao et al., 2001; Miller et al., 2014), making a network modularization analysis quite natural. Moreover, in order to design close-loop optogenetic experiments with single cell resolution, it is necessary to identify the most representative cortical ensembles that can efficiently represent different visual stimuli. The classification objective of CRFs (which strictly generalize logistic regression) provides a convenient way of defining the most representative 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 plotting the predictive power of each node in ROC space per stimulus (Figure 2C). In this way, we defined the most representative cortical ensembles as the predictive neurons that can be used to identify 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 S3).

We next investigated whether neurons from cortical ensembles identified by CRFs represent the most representative 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-3F) 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 the most representative cortical ensembles achieve the best accuracy, precision, recall and AUC when predicting the presented visual stimuli, compared with resized ensembles (Figures 3B-3F).

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 the most representative neurons. We observed that prediction performance from random groups of neurons is significantly lower than CRF ensemble performance (Figures 3G-3L), indicating that the most representative ensembles are non-random groups of neurons.

The most representative neurons from cortical ensembles identified with CRFs have similar properties and prediction performance to previously used dimensional reduction methods (Carrillo-Reid et al., 2015b; Carrillo-Reid et al., 2016) for ensemble identification (Figure 4A-4I, green and orange; Figure 4J-4M), suggesting that few neurons can be representative of cortical ensembles. On the other hand, the most representative neurons identified with CRFs are composed of a mixed population of cells with high orientation selectivity index (OSI) and cells with low orientation selectivity, while high OSI cells show reduced performance than identified ensembles (Figure 4A-4I, gray and orange; Figure 4N-Q), indicating that cortical ensembles are not purely orientation selective cells. Therefore, CRFs allow the identification of the most representative cortical ensembles that can be considered as the most significant elements for a specific experimental condition.

**Targeted manipulation of cortical microcircuits with CRFs models**

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). Therefore, a challenging issue for closed-loop optogenetic experiments to manipulate behavioral tasks in awake animals is the identification of neurons that could be used to recall cortical ensembles.

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 defined high-ranked neurons as the ones with strong node strength (Figure 5A). Indeed, single-cell two-photon optogenetic stimulation of high-ranked neurons with a strong prediction power for the optogenetic stimuli (Figures 5B and 5C) was able to evoke pattern completion of imprinted ensembles (Figure 5D), whereas low-ranked neurons with low prediction power (Figure 5B and 5C) were unable to recall imprinted ensembles (Figure 5E). Moreover, neurons with pattern completion ability appear to have stronger connections with other members of the cortical ensemble than the ones without pattern completion ability (Figure 5F and 5G). These demonstrate 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 also 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, an experimental protocol that builds new coactive ensembles (Carrillo-Reid et al., 2016). Our results demonstrated the reconfiguration of the functional connectivity between cortical ensembles after a new ensemble has been imprinted in the cortex (Figure 6A-6C). Two CRF models were constructed before and after the imprinting protocol correspondingly, while one neuron in the imprinted ensemble was being stimulated optically. After the ensemble is imprinted, the stimulated neuron shows higher the predictive power and the node strength (Figure 6B), and more connections with other members of the imprinted ensemble (Figure 6C). The density of the imprinted network also increased after population photostimulation (Figure 6D). However, the general properties of CRFs models before and after population photostimulation remain stable both within the imprinted network (Figure 6E-6H) and in the whole network (Figure S4). These suggest that the imprinted ensemble has been added to the cortical microcircuit but preserving a balance with the overall network structure.

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**

**Analysis of functional connectivity in cortical microcircuits**

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).

In the past decades, graph theory has been applied to characterize the structure and function of neuronal networks (Achard and Bullmore, 2007; Bettencourt et al., 2007; Chiang et al., 2016; Downes et al., 2012; Fair et al., 2008; Hagmann et al., 2008; Iturria-Medina et al., 2008; Oh et al., 2014; Supekar et al., 2008; Yu et al., 2008; Zuo et al., 2012). While most of these studies operated on functional recordings across multiple brain regions (Achard and Bullmore, 2007; Chiang et al., 2016; Fair et al., 2008; Hinne et al., 2013; Zuo et al., 2012), only a few have focused on the general network properties of cortical circuits with recordings from single neurons (Bonifazi et al., 2009; Sadovsky and MacLean, 2014; Stetter et al., 2012; Yatsenko et al., 2015).

The majority of methods applied to infer the functional connectivity in brain slices (Cossart et al., 2003; Ikegaya et al., 2004; Mao et al., 2001; Sadovsky and MacLean, 2014; Stetter et al., 2012) or *in vivo* (Yatsenko et al., 2015) 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. Our method provides an alternative, by directly modeling the statistical dependencies of each node.

**CRFs graph models identify 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. 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; Yu et al., 2008). The generative nature of these models restricts their ability to model arbitrary dependencies between observed 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). Therefore, CRFs appear to be promising for modeling cortical functional connectivity and for identifying the most representative 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). Thus, CRFs can be applied to datasets with thousands of interconnected neurons. However, a main constraint for applying CRFs is the number of samples in the training dataset.

Compared with the previously used descriptive methods for neuronal ensemble identification (Carrillo-Reid et al., 2015b), our approach modestly improved prediction accuracy (Figure 4). One reason could be that current CRF learning algorithms separately perform 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 during the parameter learning step can sometimes compromise the global optimality guarantees.

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 by performing singular value decomposition (SVD) (Carrillo-Reid et al., 2015a). Even though SVD can identify cortical ensembles reliably, it lacks a structured model that allows the systematic study of changes in functional connectivity.

**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). 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). 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). Scale-free networks are characterized by the existence of a small subset of nodes with high connectivity (Carrillo-Reid et al., 2015a). 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; Carrillo-Reid et al., 2016; Hagmann et al., 2008).

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 observed data. 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 added nodes are highlighted in the corresponding color. Scale bar: 50μm. (C) Relative likelihood calculated by CRFs. Gray region represents 3 S.D. Colored stripes indicate visual stimuli. Scale bar: 200 frames. (D) Temporal course of ensemble classification for horizontal (top) and vertical (bottom) drifting-gratings. Scale bar: 10 seconds. (E) ROC curve of predicting horizontal and vertical stimuli using the relative likelihood. Dashed line represent random chance. (F) Mean likelihood difference during horizontal and vertical visual stimuli. (G) Area under curve (AUC) in ROC curves for horizontal and vertical stimuli. (H) Accuracy, (I) precision and (J) recall for vertical and horizontal drifting-gratings. For conceptual model of population vector, see Figure S1; for the properties of models without added node, see Figure S2.

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

(A) Schematic representation of ensemble identification from CRF models. The activity of the neuron is set to ‘1’ or ‘0’ at each frame, and the likelihood and 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) AUC of neurons in the imaged population predicting the two visual stimuli. Ensembles are identified by taking the best AUC of each neuron, and threshold AUC with a cross-validated threshold. Examples of ensembles for horizontal (red) and vertical (blue) drifting-gratings are shown in the ROC space. (D) Spatial map of the most representative cortical ensembles from CRFs for horizontal (red, left) and vertical (blue, right) drifting-gratings. Scale bar: 50μm. For CRF methods applied in datasets with more orientations of drifting-gratings and acquired under different conditions, see Figure S3.

**Figure 3. CRF cortical ensembles are the most representative ones for visual stimuli identification**

(A) Cosine similarity between population vectors representing the most representative cortical ensembles identified with CRFs for a given visual stimuli. Each most representative 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, (D) recall, (E) AUC and (F) ROC curves of predictions from randomly down-sampled or up-sampled most representative cortical ensembles, calculated by using the similarity values to predict the time of visual stimuli. (G) Cosine similarity between population vectors of randomly sampled cells. (H) Accuracy, (I) precision, (J) recall, (K) AUC and (L) ROC curves of predictions from randomly chosen ensembles.

**Figure 4. Comparison of prediction performance with CRF, SVD ensembles and high OSI cells**

(A) Examples of core neurons identified using SVD (top, green) and CRF (all, orange), and high OSI cells (bottom, gray). Circles represent core neurons of horizontal (left) and vertical (right) visual stimuli, respectively. Neurons shared between CRF and SVD methods are represented by orange dots circled by green; neurons shared between CRF and high OSI cells are represented by orange dots circled by gray. Scale bar represents 50μm. (B) Percentage from the total population size representing core neurons for SVD and CRF methods as well as high OSI cells. (C) Percentage of shared neurons between CRF and SVD, high OSI cells. 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 with population vectors from different visual stimuli (black). (E) ROC curves of the classification result with SVD, CRF and high OSI cells. Dashed line represent random chance. (F) Area under curve (AUC) of ROC curves. (G) Accuracy, (H) precision and (I) recall of predictions for each visual stimuli using both methods. (J) Cosine similarity between population vectors representing the most representative cortical ensembles identified with SVD for a given visual stimulus. Each SVD ensemble was randomly down-sampled (dark green). The cosine similarity of population vectors belonging to different visual stimuli is shown in light green. (K) Accuracy, (L) precision and (M) recall of predictions from randomly down-sampled SVD ensembles. (N) Cosine similarity between population vectors of high OSI cells. (O) Accuracy, (P) precision and (Q) recall of predictions from randomly down-sampled high OSI cells.

**Figure 5. 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 node strength 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 color tone represents the strength of ; node color represents the rank of cells. Scale bar: 50μm. (C) Node strengths and the AUC values of neurons in the imprinted ensemble for the single cell photostimulation. AUC values were calculated by predicting the photostimulation time of the recalling neuron using the cosine similarity of population vector of each individual neuron in the imprinted ensemble. Neuron in red represents the same high-ranked neuron highlighted in red in (B); neuron in blue represents the same low-ranked neuron highlighted in blue in (B). Dashed line represents the AUC of the imprinted ensemble predicting the photostimulation time of the recalling neuron. (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: 50μm. (F) Neuron with pattern completion ability (red, lower right corner) has stronger connections with other members in the imprinted ensemble. (G) Neuron without pattern completion ability (blue, lower right corner) has weak connections with other members in the imprinted ensemble.

**Figure 6. Reconfiguration of cortical microcircuits induced by two-photon optogenetic population manipulation**

(A) CRF graphical models trained with data 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 (10 trials). Edge color tone represents the strength of ; node color represents the node strength. Node size represents the node degree. Scale bar: 50μm. (B) Node strength and classification AUC of the stimulated neuron for single cell stimulation before (left) and after (right) photostimulation trials. Dashed lines represent the AUC of the imprinted ensemble predicting the time of single cell photostimulation. (C) The connection density of the stimulated neuron with other members of the imprinted ensemble before (left) and after (right) photostimulation trials. Red dot represents the stimulated neuron. (D) Graph density, (E) node strength, (F) node degree, (G) clustering coefficient and (H) centrality values of the imprinted ensemble before (black) and after (blue) imprinting protocol remained stable indicating that the imprinted ensemble reconfigured the network structure preserving the global network properties (n = 74 neurons; Wilcoxon signed rank-test ; n.s. p>0.05). Global properties of the whole network remain unchanged too (Figure S4).

**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.

Cossart, R., Aronov, D., and Yuste, R. (2003). Attractor dynamics of network UP states in the neocortex.

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.

Ikegaya, Y., G., A., R., C., D., A., I., L., D., F., and Yuste, R. (2004). Synfire chains and cortical songs: temporal modules of cortical activity.

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.

Mao, B.Q., F., H.-S., Aronov, D., Froemke, R.C., and Yuste, R. (2001). Dynamics of spontaneous activity in neocortical slices.

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.

**STAR methods**

**KEY RESOURCES TABLE**

(Need a separate file for key resources table)

Code for training CRF models can be found at <https://github.com/kuitang/fwmatch-public>.

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Luis Carrillo-Reid ([lc2998@columbia.edu](mailto:lc2998@columbia.edu)).

**EXPERIMENTAL MODELS AND SUBJECT DETAILS**

None.

**METHOD DETAILS**

**Spike detection of Allen Institute Brain Observatory dataset**

Calcium traces and dF/F were obtained using the SDK provided by Allen Institute. Then, dF/F values of the top and bottom 30% were removed to obtain a baseline activity, and a threshold was defined by the mean of the baseline plus 5 times standard deviation of baseline values. Spikes were detected as time points where dF/F values are higher than the threshold.

**Conditional Random Fields**

We construct a conditional random field (CRF) as previously published (Tang et al., 2016), using indicator feature vector where , for each edge and node, 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). 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 learned a graph structure using -regularized neighborhood-based logistic regression (Ravikumar et al., 2010):

,

where

.

Here is a regularization parameter that controls the density of constructed structure. Then, a graph structure is learnt by thresholding the edge potentials with a given density preference . Edges with potential values within top quantile were kept as the final structure. It is worth noting that although could bias the result, varying does not lead to density values that differ much. This is probably because of the sparse nature of the obtained Ising model.

Based on the learned 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):

Here is a regularization that controls the learnt parameters. Cross-validation was done to find the best , and via model likelihood. We varied with 6 values between 0.002 and 0.5, d with 6 values between 0.25 and 0.3, and with 5 values between 10 and 10000, all sampled uniformly. To obtain the best model parameters, 90% data were used for training, while 10% data were withheld for cross-validation. The best model parameters were determined by calculating the likelihood of the withheld data and selecting the parameter set with a locally maximum likelihood in the parameter space.

**Identifying the most representative cortical ensembles**

To find the most representative 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 computed the likelihood difference vector

and calculated the standard receiver operating characteristic (ROC) curve with the ground truth as the timing of each presented visual stimuli. The prediction ability of all nodes for all presented stimuli is then represented by an area under curve (AUC) matrix , where represents the AUC value of node predicting stimulus .

The most representative cortical ensemble for stimulus is defined by the following procedure: (1) find the nodes that has maximum value at column ; (2) vary a threshold between 0.5 and 0.9 with a step size of 0.05, and take all the nodes in (1) that has AUC values larger than the threshold; (3) take the population vector of the resulting nodes, calculate the cosine similarity with the binary data, predict timing of stimulus from the cosine similarity (see next session for details), and calculate accuracy of prediction; (4) take the threshold with the best prediction accuracy; (5) the final ensemble is define using step (3), using the threshold from (4).

**Prediction with cosine similarity**

Identified most representative 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 population vectors in a high-dimensional space.

**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 the node .

**QUANTIFICATION AND STATISTICAL ANALYSIS**

CRF models were trained using the Columbia Yeti shared HPC cluster. MATLAB R2016a (MathWorks) were used for data analysis. Statistical details of each specific experiments, including the statistical methods, the meaning and value of n, and the significance level can be found in figure legends.

**DATA AND SOFTWARE AVAILABILITY**

The analysis of simultaneous two-photon imaging and two-photon optogenetic stimulation data was performed from experiments previously published (Carrillo-Reid et al., 2016).

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. The experiments IDs are: 511507650, 511509529, 511510650, 511510670, 511510718, and 511510855.

All data processing, analysis and plotting code in this paper can be found at <https://github.com/hanshuting/graph_ensemble>.

**Supplementary Information**

**Supplementary Figure S1. 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.

**Supplementary Figure S2. Added nodes does not affect graph properties**

(A) CRF graphs of baseline model (no added nodes) and the added node model, trained with the same experiment. Edge color represents ; node color represents node strength. Node size represents the node degree. Scale bar represents 50μm. (B) Graph densities (n=6 experiments), (C) node strength (one experiment shown here; 3/6 experiments show significant change), (D) node degrees (one experiment shown here; 6/6 experiments show significant change), (E) clustering coefficients (one experiment shown here; 6/6 experiments show significant change) and (F) centrality (one experiment shown here; 2/6 experiments show significant change) comparison between the two models (Wilcoxon signed rank-test; n.s. p>0.05).

**Supplementary Figure S3. CRF ensembles are able to predict multiple stimuli**

(A) An example of constructed CRF graph from the Allen Brain Observatory dataset, with four orientations of drifting grating stimuli (squares). Edge color indicates the strength of inferred connections; node size indicates the node degrees. (B) Temporal course of ensemble classification for four drifting-gratings. Colored stripes indicate visual stimuli. Scale bar represents 200 frames. (C) Accuracy, (D) precision and (E) recall of prediction with CRF model trained with temporal frequency 1Hz, on dataset of temporal frequency 2, 4, 8, and 15Hz (n=6 mice).

**Supplementary Figure S4. Global network properties of cortical microcircuits induced by two-photon optogenetic population manipulation**

(A) Density, (B) node strength, (C) node degree, (D) clustering coefficients and (E) centrality values of the whole network remain unchanged after the imprinting protocol. (n = 74 neurons; Wilcoxon signed rank-test; n.s. p>0.05).