Infer Functional Connectivity from Partially Observed Neural Signals

Xincheng Tan
Harvard University
Cambridge, MA
xinchengtan@g.harvard.edu

Xiaomin Li
Harvard University
Cambridge, MA
xiaominli@g.harvard.edu

Ce Luo
Massachusetts Institute of Technology
Cambridge, MA
celuo@mit.edu

***Abstract*—Functional connectivity is imperative to understanding how the neurons interact with each other. However, due to technology or budget limitations, researchers may not be able to obtain the simultaneous recording of all neuron pairs in their experiments, and some neurons are never observed at the same time. In this study, we investigated a data-driven approach to infer functional connectivity when only a subset of neurons are simultaneously recorded. Our method is able to obtain ???**

***Index Terms*—functional connectivity, graphical model, matrix completion, computational neuroscience**

I. Introduction

The connectivity pattern among neuron populations has been a fundamental topic in understanding the brain. There are two major types of connectivity: structural and functional connectivity. Structural connectivity is defined as the existence of the white matter tracts physically interconnecting the brain regions, while functional connectivity is the statistical correlation between neural signals, such as fMRI signals [13]. Although structural connectivity provides anatomical basis for how the neurons may interact with each other, it does not necessarily contain functional connectivity – researchers have discovered that certain brain regions without physical connections can exhibit strong functional connectivity [13]. Moreover, functional connectivity in fact provides extra insight into how neurons function together. For example, studies have found a strong correlation between static functional connectivity and certain neurological diseases such as depression, schizophrenia, and Alzheimer's disease [8].

Recently, with the development in functional neuroimaging, researchers are able to quantify and record the activities of a large neuron population in various animals. In particular, two-photon calcium imaging technology enables non-invasive, simultaneous monitoring of thousands of neurons in mice. To derive meaningful insights from these video datasets, a powerful calcium imaging analysis pipeline, which includes neuron detection, motion correction, spike inference etc., has been developed to extract more computationally tractable neural signals [10].

However, the cost of acquiring and maintaining such technology is prohibitively high and more affordable options may only be able to record a subset of neurons simultaneously, where some neuron pairs are never observed at the same time. These scenarios are especially challenging to estimate functional connectivity with off-the-shelf statistical models which require full observations. In this study, we use graphical lasso as the primary framework for estimating the sparse functional connectivity [5].

II. Related Works

A. Graphical Models

Graphical models combine graph theory and probability theory to infer complex network structure within large systems. They aim to uncover an underlying graph $G(V, E)$, where each node represents a feature of interest and each edge reflects the conditional dependence between the two features. To estimate the functional connectivity, we focus on the existence of the connection instead of the direction of such connection. Thus, we only discuss undirected graphical models for the rest of this paper.

Formally, let $X = (X_v : v \in V = \{1, 2, ..., p\})$ be a p-dimensional random vector indexed by the vertices of an undirected graph $G = (V, E)$ where $E \subseteq V \times V$. Moreover, we say X satisfies the *pairwise Markov property* with respect to G if:

$$X_i \perp X_j | X_{V \setminus \{i,j\}} \iff (i, j) \notin E \qquad \forall i, j \in V, i \neq j$$

The pairwise Markov property translates each absence of an edge into 'full' conditional independence. In other words, two random variables are conditionally independent if and only if the nodes representing them are not connected under the graphical model representation.

B. Gaussian Graphical Lasso

A graphical model is Gaussian when $X = (X_1, ..., X_p)$ follows a Gaussian distribution $\mathcal{N} = (\mu, \Sigma)$. A Gaussian conditional dependence graph can be estimated by determining the zero entries of the inverse covariance matrix $\Theta = \Sigma^{-1}$, also referred to as *precision* matrix. That is, if $\Theta_{ij} = 0$, $(i, j) \notin E$. Therefore, to estimate the underlying conditional dependence graph, we need to estimate the inverse covariance matrix of the underlying data distribution [3].

Gaussian graphical model (GGM) typically uses the maximum likelihood estimator (MLE) to estimate Θ. Let S be the

sample covariance matrix for an i.i.d. Gaussian sample of size n. The MLE maximizes the following log-likelihood function

$$L_{ggm}(\Theta) = \log\det(\Theta) - \mathrm{tr}(S\Theta) \quad (1)$$

C. Graphical Lasso

A popular variant of general undirected Gaussian graphical model is the *graphical lasso* (Glasso) proposed by Yuan & Lin in [14] and Banerjee *et al.* in [1]. Based on the general Gaussian graphical model, it adds an ℓ_1 regularization term in the MLE objective function:

$$L_{GLasso}(\Theta) = \log\det(\Theta) - \mathrm{tr}(S\Theta) - \lambda||\Theta||_1 \quad (2)$$

where λ is a regularization parameter that controls the sparsity of the graph estimation.

The formulation of GLasso has a few advantages. First, after adding the regularization term, its objective function is still convex and the convex optimization can be solved in polynomial time. Second, when the number of observations is insufficient ($n < p$), GLasso has strong statistical guarantees which include convergence in Frobenius and spectral norm [7]. Third, GLasso provides a hyperparameter λ to control the sparsity in the graph while preserving the statistical significance of the selected connections, which is very convenient if we are only interested in certain strong connections and wish to ignore the weaker ones.

D. Missing Data

A few existing methods have been investigated to handle missing observations [4], [9], [12]. However, these methods only apply to partial data matrix that does not lead to an incomplete sample covariance matrix. Moreover, they assume the observations are missing at random (MAR), which does not hold in our scenario.

Another direction is to solve the matrix completion problem by recovering the sample covariance. In [2], Bishop and Yu propose a symmetric positive semidefinite matrix completion method by stitching in order the rank-factorized components of each revealed principle submatrices. However, their method imposes hidden constraint on the matrix rank and fails to recover the PSD matrix when the rank of the full matrix is high. In our dataset, the true sample covariance matrix is almost always positive definite, and thus a low-rank approximation is unlikely to be appropriate.

III. Proposed Method

A. Data-driven Covariance Completion

In this study, we propose a data-driven approach to estimate functional connectivity by leveraging the auxiliary data to recover the incomplete sample covariance matrix. Note that even though certain neuron pairs do not have simultaneous observations, the behavioral features, such as running speed and pupil area, are recorded throughout the entire experiment session. We hypothesize that neural functional connectivity strongly associates with how similar each pair of neuron act in accordance with certain behavioral events, and in identifying such an association, we are able to recover their connectivity.

Denote $X \in R^{N\times T}$ as the neural signal recordings where N is the number of neurons and T is the number of time points. Certain blocks of X are missing such that its sample covariance matrix S_{obs} is incomplete. Denote $P \in R^{p\times(N\times N)}$ as pairwise similarity matrices over p different variables without missing entries.

We propose to learn a function f that maps each $P_{i,j} \in R^p$ to $(S_{obs})_{i,j} \neq \emptyset$ and use the learned f and P to infer the missing entries in S_{obs}. If the recovered sample covariance $\hat{S}$ is not positive semidefinite, we use existing solutions for nearest correlation matrix (NCM) to correct $\hat{S}$ to its closest positive semidefinite matrix [6]. Then apply graphical lasso on $\hat{S}$ with the regularization parameter λ to estimate the precision matrix $\hat{\Theta}$. By definition, neuron i and j are functionally connected if and only if $\hat{\Theta}_{i,j} \neq 0$. The pseudocode of our method is summarized in Algorithm [1] below.

Algorithm 1 Auxiliary-feature-aided Graphical Lasso

Input: Incomplete data matrix $X \in R^{N\times T}$, regularization strength λ, pairwise feature matrix $P \in R^{p\times(N\times N)}$, predictive model f

Output: Precision matrix $\hat{\Theta}$ and covariance matrix $\hat{\Sigma}$

```
S_obs ← incomplete sample covariance of X;
Fit f which maps P to S_obs entry-wise;
Ŝ ← S_obs + f(P)_obs ;
if Ŝ is not positive semidefinite then
    Perturb Ŝ to the nearest PSD matrix;
Σ̂, Θ̂ ← Graphical lasso(Ŝ, λ);
Return Σ̂, Θ̂;
```

B. Pairwise Feature Matrices

In this study, the pairwise feature matrices are constructed from two measures: physiological proximity and tuning curve similarity. For the former, we choose the Euclidean distance since the neuron position is in 3D coordinate and is low-dimensional. For the latter, we apply the kernel methods which encode a high-dimensional feature space and may potentially enhance the predictive power of f. Specifically, we experiment with three kernels: linear kernel, Cosine similarity and Gaussian kernel. Note that in our implementation, linear kernel corresponds to the Pearson correlation of the turning curves.

IV. Experiments

A. Datasets

The dataset in this study is published by Stringer which includes nine simultaneous calcium imaging recordings of the mice primary visual cortex [11]. Each recording captures over ten thousand neurons over 9 to 11 planes equally spaced with 35 μm in depth. The recordings range from 90 to 120 minutes with a recording frequency of either 2.5 or 3 Hz. The auxiliary variables include nucleus 3D location, pupil area, running speed, and preprocessed facial and whisker image SVD.

B. Setup and Procedure

In our experiment, we use a subset of neuron cubes in four consecutive neuron layers in five sessions: M150824_MP019_2016-04-05, M160907_MP028_2016-09-26, M161025_MP030_2016-11-20, M161025_MP030_2017-06-16, M161025_MP030_2017-06-23. Each selected subset contains 200 to 400 neurons and we use a subset of 30 consecutive minutes. We first fit a GLasso model on the full observations for each recording and each corresponding λ is selected by a 4-fold cross validation. Note that when computing the sample covariance matrix, we in fact compute the sample correlation matrix which is normalized in mean and variance.

Then, we simulate the missing observations by neuron plane. As in Fig 1, only neurons in the adjacent two layers have simultaneous observations. We assume the imaging tool scans the cortex circularly from cortex surface to bottom and stays at every two adjacent layers for around 10 seconds.

Next, we fit a predictive model f on the pairwise feature matrices. We leverage two types of predictive models: linear regression and ridge regression. In the ridge regression fitting, the regularization strength is automatically selected based on a 5-fold cross validation.

To construct the predictors of f, we use the tuning curve similarity of the following behavioral features: running speed, change in running speed, pupil area and change in pupil area. Since these variables are all discrete, we bucket their values between their respective range with arbitrary number of bins. Details can be found in our GitHub repository: https://github.com/XinchengTan/Recover_Functional_Connectivity/.

After f is learned on the pairwise feature matrices for all the observed pairs, we apply it to infer the missing entries in sample covariance. In all our experiments, the resulting $Shat$ is always non PSD, so we use the alternating projection method developed in [6] to perturb it to the nearest PSD matrix.

At last, we apply Glasso with a preset λ on $\hat{S}$ to estimate the precision matrix. The connectivity graph is correspondingly derived based on its non-zero entries.

C. Evaluation Metrics

We evaluate our method based on two aspects: the quality of the covariance/precision matrix and the similarity of the connectivity pattern. For the first, we examine

1. Rooted mean squared error (RMSE) in sample covariance recovery and downstream precision matrix estimation: $||S_{full} - \hat{S}||_{frob}/N$ and $||\Theta_{full} - \hat{\Theta}||_{frob}/N$
2. Off-diagonal Pearson correlation of S_{full} and $\hat{S}$, and Θ_{full} and $\hat{\Theta}$.

For the second, we evaluate the accuracy, precision, recall and F1 score of the detected edges with respect to the connectivity under the full data. Since the connectivity graph is sparse, accuracy is not very indicative and we focus on F1 score as our primary performance indicator.

The results below are averaged across the five selected sessions.

D. Results

1) Auxiliary Predictors: We examine how different auxiliary variables in the pairwise feature matrices influence the connectivity estimation. From Fig 2, we do not see a significant difference in their predictive power in recovering the final connectivity graph. However, when all five variables are used, the Gaussian-kernelized similarity matrices (red bar) demonstrate the best average F-1 score compared with linear kernel and cosine similarity (pink and purple bar). The

2) Similarity Measure: The RMSE and correlation in sample covariance estimation show a quite different picture. In Fig 3 and Fig 4, it is obvious that the pairwise matrices based on Gaussian kernel lead to the lowest RMSE and highest off-diagonal correlation. This phenomenon applies to any individual tuning variable as well as all five together.

E. Regularization Strength

Another important parameter that influences the connectivity estimation is the regularization strength. For all five sessions, the optimal λ on the full recording is 0.1 selected by cross validation.

As we adjust λ in the graphical lasso for partial observations, the F1 score first increases and then decreases, reaching a sweet spot at $\lambda \approx 0.13$ for all five sessions. It is slightly higher than the optimal value under full data because the inferred sample covariance matrix is not perfect and introduces noise. Therefore, it requires a stronger regularization strength to preserve the same level of sparsity. However, if λ is too large, the model suffers from a high miss rate in that the estimated connectivity is overly sparse and fails to detect a lot of edges. On the other hand, when λ is too small, the estimated graph is too dense and false discovery rate is very high.

V. Future Work

This study only digs an iceberg of the possible techniques to recover functional connectivity under partial observations. The future work lies in the following aspects.

First, it would be ideal to automate λ selection under partial data without knowing the optimal parameter for the full data. To achieve this, we can fit graphical lasso under multiple subsets of neurons with simultaneous observations. Cross validating on all subsets, we can obtain a list of optimal λ for each subset and ensemble them to obtain a proper value for graphical lasso on the partial data.

Second, the predictive models used in our experiments are quite simplistic as it aims to find a uniform, linear combination of the pairwise feature matrices to approximate the sample covariance matrix. More sophisticated methods, such SVM and ensemble learning, should be investigated to improve its predictive power.

In fact, we experimented using neural networks to learn the precision matrix with simulated missing observations. We were interested in whether conditional dependency, which

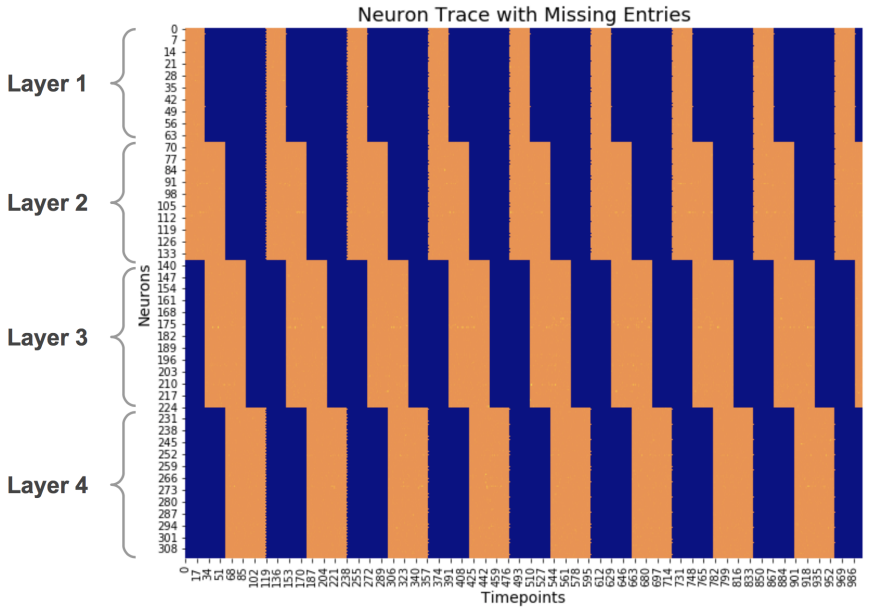

Fig. 1. Example of a partially observed data matrix. The orange blocks represent observed neural activities while the dark blue blocks are missing entries. Some neuron layers, e.g. layer 1 and 4, are never simultaneously observed.

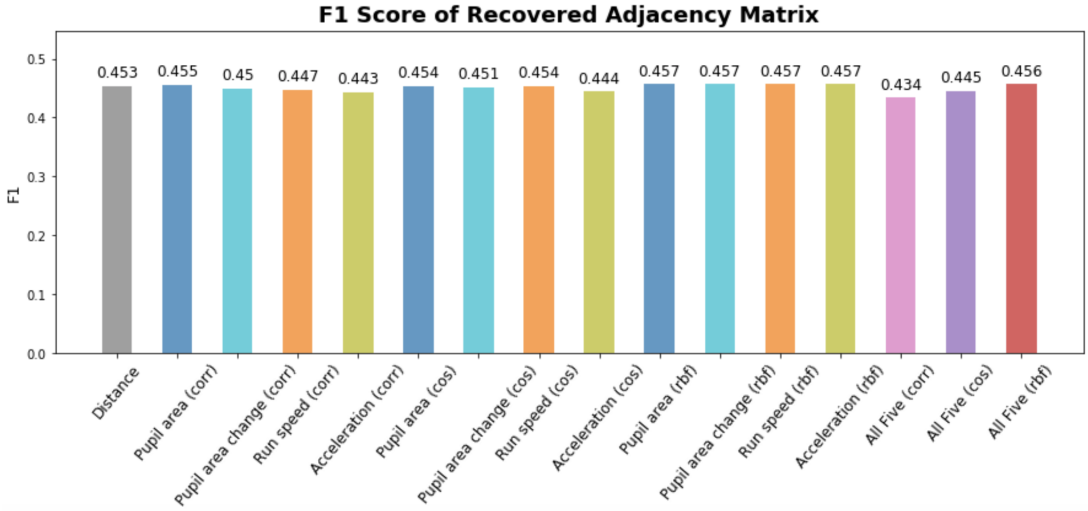

Fig. 2. Connectivity F1 Score under different feature predictors of three kernels

represents functional connectivity of a pair of neurons, can be learned directly by neural networks with ReLu activation, but the current RMSE of this approach is about $\sqrt{16} = 4$.

Third, our current study only compares three similarity kernels. It would be worthwhile to experiment with other kernels, such as polynomial kernel and Fourier kernel, to determine which one fits best.

Last, the current recordings are all from mice in a spontaneous status. There are also neuron recordings when different types of visual stimuli are presented. It would be interesting to see whether the tuning to the visual stimuli would help the connectivity estimation.

ACKNOWLEDGMENT

We acknowledge Bahareh Tolooshams and Prof. Demba Ba for the helpful discussion and suggestions.

REFERENCES

[1] O. Banerjee, L. E. Ghaoui, and A. d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, 9(15):485–516, 2008.

[2] W. E. Bishop and B. M. Yu. Deterministic symmetric positive semidefinite matrix completion. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[3] M. Drton and M. H. Maathuis. Structure learning in graphical modeling, 2016.

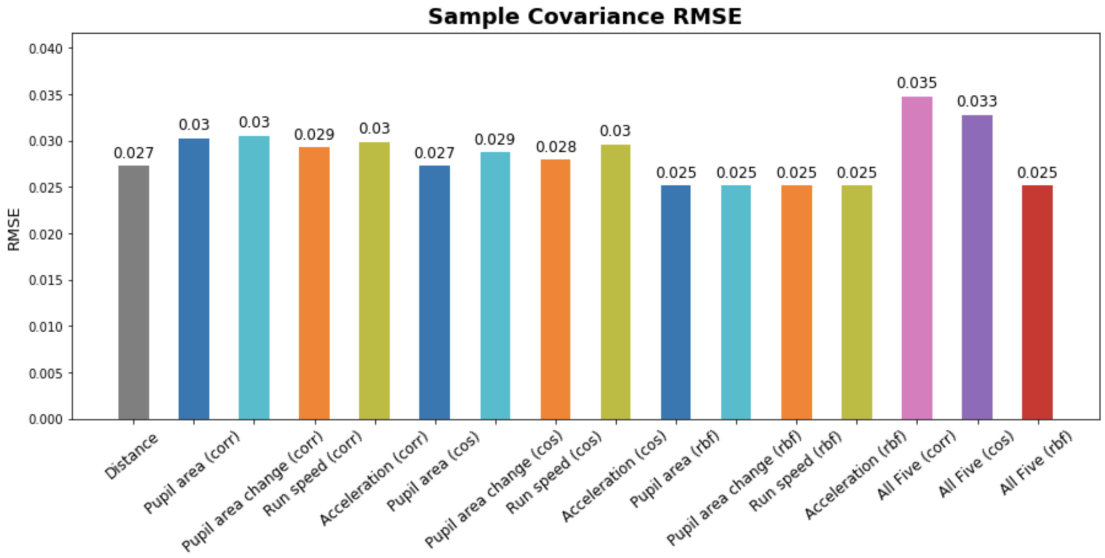

Fig. 3. RMSE of $S, \hat{S}$ under different feature predictors of three kernels

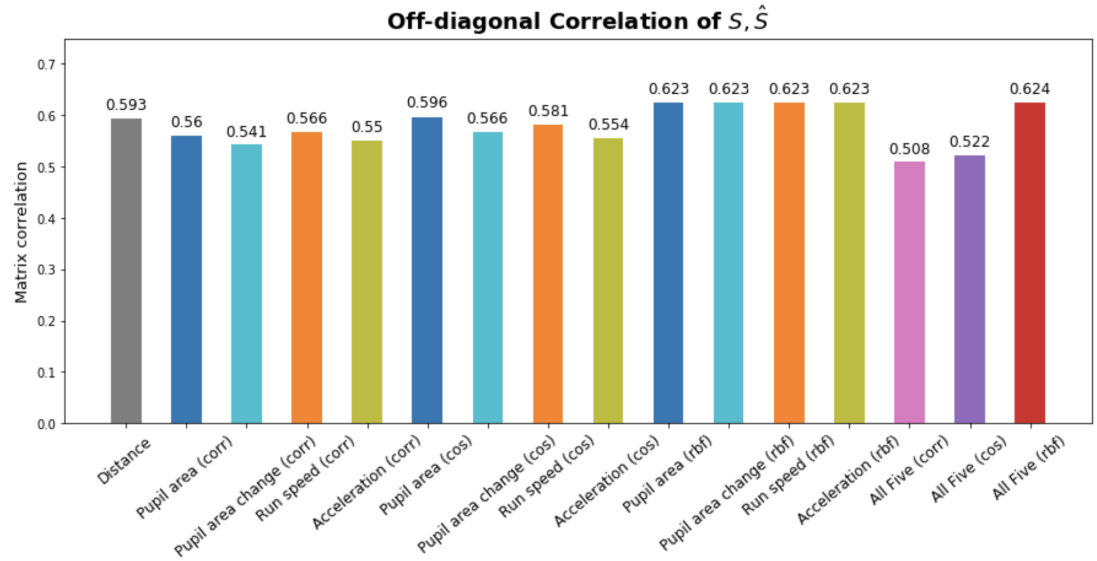

Fig. 4. Correlation of $S, \hat{S}$ under different feature predictors of three kernels

[4] R. Fan, B. Jang, Y. Sun, and S. Zhou. Precision matrix estimation with noisy and missing data, 2019.

[5] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the lasso, 2007.

[6] N. J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002.

[7] C.-J. Hsieh, M. A. Sustik, I. S. Dhillon, and P. Ravikumar. Quic: Quadratic approximation for sparse inverse covariance estimation. *J. Mach. Learn. Res.*, 15(1):2911–2947, Jan. 2014.

[8] M. M. Jones DT, Vemuri P. Non-stationarity in the resting brain's modular architecture. *PLos One*, 7, 2012.

[9] M. Kolar and E. Xing. Estimating sparse precision matrices from data with missing values. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, 1, 01 2012.

[10] E. A. Pnevmatikakis. Analysis pipelines for calcium imaging data. *Current Opinion in Neurobiology*, 55:15 – 21, 2019. Machine Learning, Big Data, and Neuroscience.

[11] C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, and K. D. Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437), 2019.

[12] N. Städler and P. Bühlmann. Missing values: sparse inverse covariance estimation and an extension to sparse regression. *Statistics and Computing*, 22(1):219–235, Dec 2010.

[13] L. Q. Uddin. Complex relationships between structural and functional brain connectivity. *Trends in Cognitive Sciences*, 17:600 – 602, 2013.

[14] M. Yuan and Y. Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

APPENDIX

This is appendix.

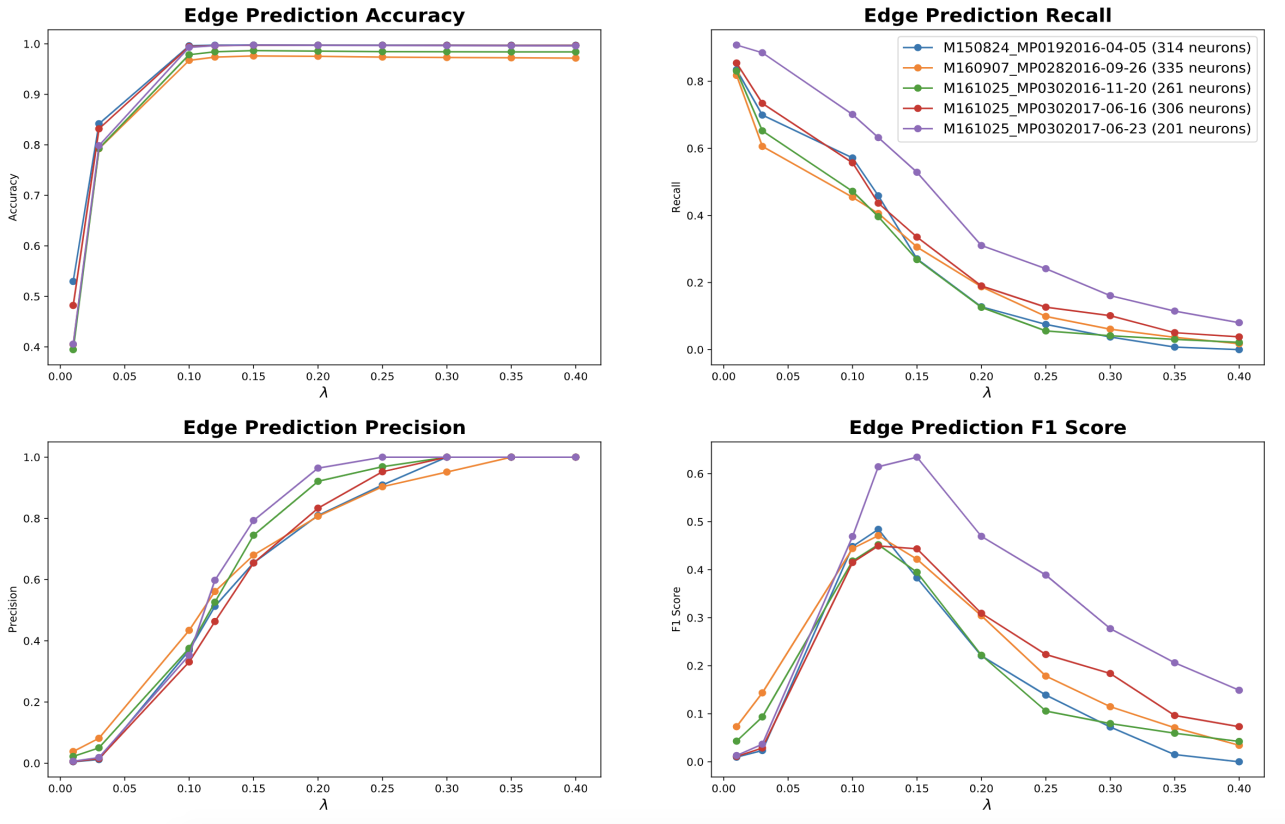

Fig. 5. Quality of Connectivity Estimation over Different λ