Simple connectome inference from partial correlation statistics in calcium imaging

Antonio Sutera, Arnaud Joly, Vincent François-Lavet, Zixiao Aaron Qiu, Gilles Louppe, Damien Ernst and Pierre Geurts

a.sutera@ulg.ac.be Code source available at https://github.com/asutera/kaggle-connectomics

Abstract

In this work, we propose **a simple yet effective solution** to the problem of connectome inference in calcium imaging data. The proposed algorithm consists of **two steps :**

i) processing the raw signals to detect neural peak activities,

ii) inferring the degree of association between neurons from partial correlation statistics.

We summarise the methodology that led us to win the Connectomics Challenge, propose a simplified version of our method, and finally compare our results with respect to other inference methods.

Context

Without being perfect, **calcium imaging** currently allows for **real-time and simultaneous observation of neuron activity** from thousands of neurons, producing **individual time-series representing their fluorescence intensity**. From these data, the connectome inference problem amounts to retrieving the synaptic connections between neurons on the basis of the fluorescence time-series. This problem is difficult to solve because of **experimental issues, including**

i) masking effects (i.e., some of the neurons are not observed or confounded with others),

ii) the low sampling rate of the optical device with respect to the neural activity speed, or

iii) the slow decay of fluorescence.

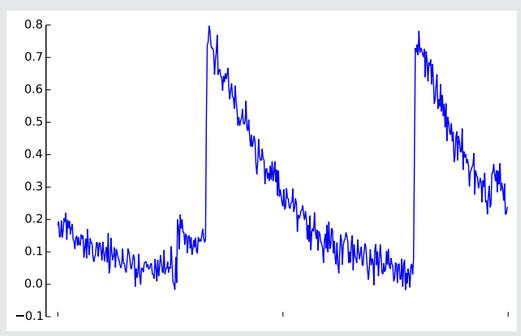
Signal processing

Under the simplifying assumption that neurons are on-off units, characterised by short periods of intense activity, or peaks, and longer periods of inactivity, the first part of our algorithm consists of cleaning the raw fluorescence data. More specifically, time-series are processed using standard signal processing filters in order to :

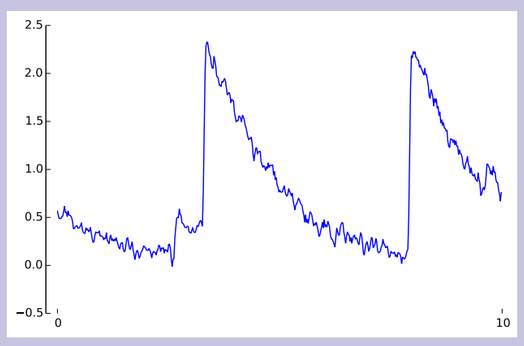
i) remove noise mainly due to fluctuations independent of calcium, calcium fluctuations independent of spiking activity, calcium fluctuations in nearby tissues that have been mistakenly captured, or simply by the imaging process ;

ii) to account for fluorescence low decay ; and

iii) to reduce the importance of high global activity in the network.

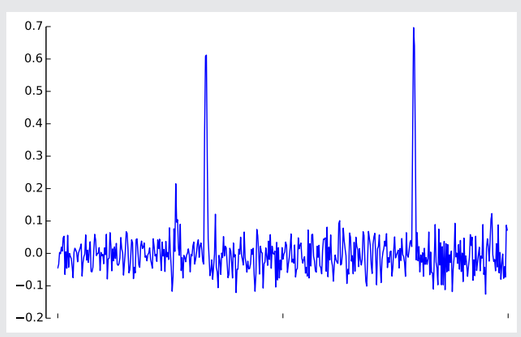


Original curve



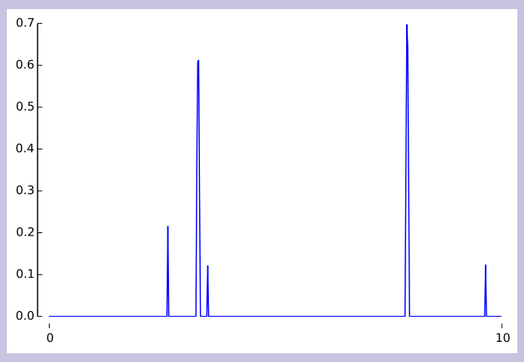
Applying a low-pass filter

$$f_1(x_i^t) = x_i^{t-1} + x_i^t + x_i^{t+1},$$
$$f_2(x_i^t) = 0.4x_i^{t-3} + 0.8x_i^{t-2} + x_i^{t-1} + x_i^t.$$



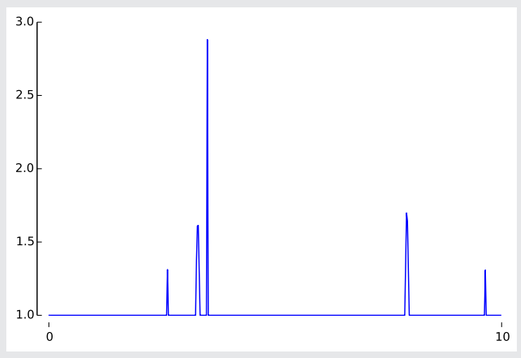
Applying a high-pass filter

$$g(x_i^t) = x_i^t - x_i^{t-1},$$



Applying a hard-threshold filter

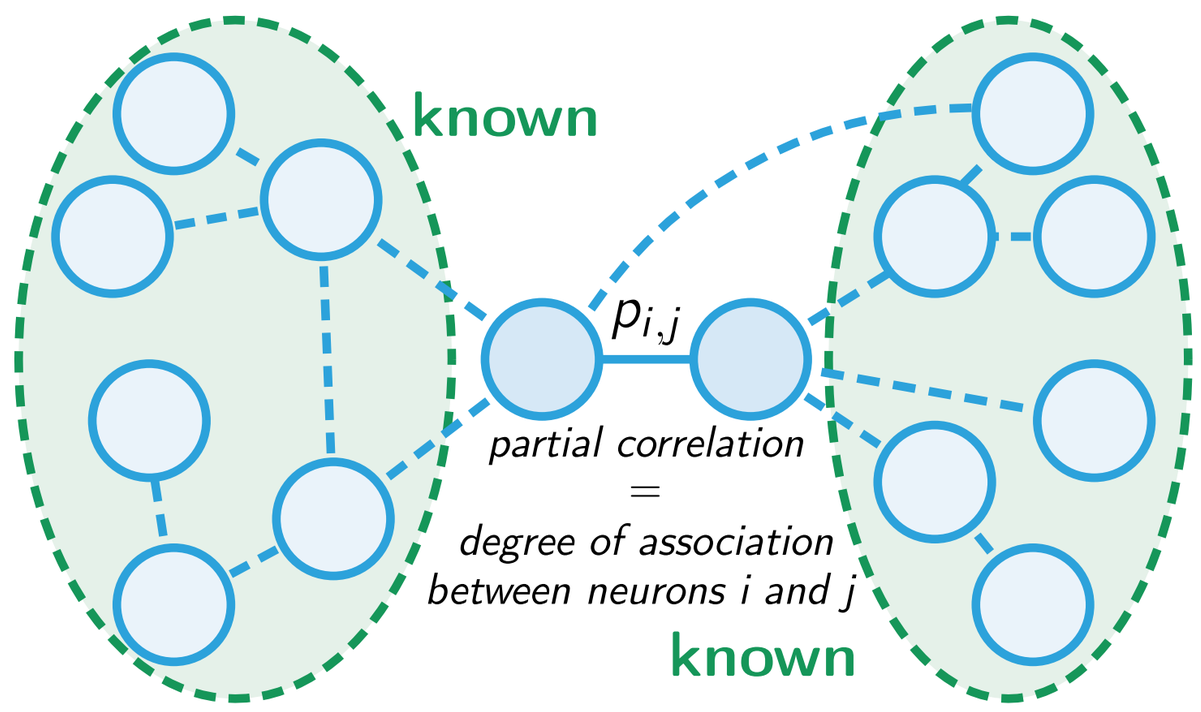
$$h(x_i^t) = x_i^t \mathbb{1}(x_i^t \geq \tau) \text{ with } \tau > 0,$$



Magnify the importance of spikes that occur in cases of low global activity

$$w(x_i^t) = (x_i^t + 1)^{1 + \frac{1}{\sum_j x_j^t}},$$

Connectome inference



Given the assumption of random variables drawn from the same time-invariant joint probability distribution, we then propose to use as **a measure of the degree of association of the connection between two neurons** i and j, the ***Partial correlation*** coefficient $p_{i,j}$ between their corresponding random variables X_i and X_j, defined by :

$$p_{i,j} = -\frac{\Sigma^{-1}_{ij}}{\sqrt{\Sigma^{-1}_{ii}\Sigma^{-1}_{jj}}},$$

where Σ^{-1}, known as the precision or concentration matrix, is the inverse of the covariance matrix Σ of X.

Assuming that the distribution P_X is a multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, it can be shown that $p_{i,j}$ is zero if and only if X_i and X_j are independent given all other variables in X, i.e., $X_i \perp X_j | X^{-i,j}$ where $X^{-i,j} = X \setminus \{X_i, X_j\}$.

Advantages and drawbacks :

✓ Only **direct** associations

✓ Filter out spurious **indirect** effects

✓ **Symmetric** (i.e. $p_{i,j} = p_{j,i}$)

✗ Edge orientation

✗ Sensitive to the value of parameters in the filtering process

Experiments

These results **clearly** show the importance of the **filters and PCA** (especially for AUPRC). **Taking the average over various parameter settings** gives an improvement of 10% in AUPRC but only a minor change in AUROC. The last row ("Full method") shows the final performance of the **method specifically tuned for the challenge** (see paper for all details). Although this tuning was decisive to obtain the best performance in the challenge, it does not significantly improve either AUROC or AUPRC.

	Performance with partial correlation							
	AUROC				AUPRC			
with *Method* \ on *normal-*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
No filtering	0.777	0.767	0.772	0.774	0.070	0.064	0.068	0.072
$h \circ g \circ f_1$	0.923	0.925	0.923	0.922	0.311	0.315	0.313	0.304
$w \circ h \circ g \circ f_1$	0.931	0.929	0.928	0.926	0.326	0.323	0.319	0.303
+ PCA	0.932	0.930	0.928	0.926	0.355	0.353	0.350	0.333
Averaging	0.937	0.935	0.935	0.931	0.391	0.390	0.385	0.375
Full method	**0.943**	**0.942**	**0.942**	**0.939**	**0.403**	**0.404**	**0.398**	**0.388**
PC	0.886	0.884	0.891	0.877	0.153	0.145	0.170	0.132
GTE	0.890	0.893	0.894	0.873	0.171	0.174	0.197	0.142
GENIE3	0.892	0.891	0.887	0.887	0.232	0.221	0.237	0.215

Partial correlation and averaged partial correlation clearly outperform all other methods on all datasets. The improvement is more important in terms of AUPRC than in terms of AUROC. As expected, Pearson correlation performs very poorly in terms of AUPRC. GTE and GENIE3 work much better, but **these two methods are nevertheless clearly below partial correlation**.

Conclusion

We outlined a simple but efficient methodology for the problem of connectome inference from calcium imaging data consisting in processing the raw signals and then inferring the degree of association between neurons from partial correlation statistics. Its **simplified variant outperforms other network inference methods** while **its optimized version proved to be the best method on the Connectomics Challenge**. Given its simplicity and good performance, we therefore believe that the methodology presented in this work would constitute a solid and easily-reproducible baseline for further work in the field of connectome inference.