



**HAL**  
open science

## Reconstructing the functional connectivity of multiple spike trains using Hawkes models

Régis C. Lambert, Christine Tuleau-Malot, Thomas Bessaih, Vincent Rivoirard, Yann Bouret, Nathalie Leresche, Patricia Reynaud-Bouret

► **To cite this version:**

Régis C. Lambert, Christine Tuleau-Malot, Thomas Bessaih, Vincent Rivoirard, Yann Bouret, et al.. Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *Journal of Neuroscience Methods*, 2018, 297, pp.9-21. hal-01740126

**HAL Id: hal-01740126**

**<https://hal.science/hal-01740126v1>**

Submitted on 21 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Reconstructing the functional connectivity of multiple spike trains using Hawkes models

Régis C. Lambert<sup>a,\*</sup>, Christine Tuleau-Malot<sup>b</sup>, Thomas Bessaih<sup>a</sup>,  
Vincent Rivoirard<sup>c</sup>, Yann Bouret<sup>d</sup>, Nathalie Leresche<sup>a</sup>,  
Patricia Reynaud-Bouret<sup>b,\*</sup>

March 21, 2018

<sup>a</sup>Sorbonne Universités, UPMC Univ Paris 06, INSERM, CNRS, Neurosciences Paris Seine - Institut de Biologie Paris Seine (NPS - IBPS), 75005 Paris, France

<sup>b</sup> Université Côte d'Azur, CNRS, LJAD, France.

<sup>c</sup> Ceremade, CNRS, UMR 7534, Université Paris-Dauphine, PSL Research University, 75016 Paris, France

<sup>d</sup> Université Côte d'Azur, CNRS, INΦNI, France.

\* Corresponding author :

R.C. Lambert

Université Pierre et Marie Curie- Neuroscience Paris Seine (NPS)

case 16 BatB 5eme et.

9, quai st Bernard 75252 PARIS CEDEX 05

phone : 33 1 44 27 25 83 ; fax : 33 1 44 27 25 84 ; regis.lambert@upmc.fr

P. Reynaud-Bouret

Laboratoire de Mathématiques J.A. Dieudonné

UMR n° 7351 CNRS UNS

Université de Nice-Sophia Antipolis

Parc Valrose, 06108 Nice Cedex 02, France

phone : 33 4 92 07 60 33 ; fax : 33 4 93 51 79 74 ; reynaudb@unice.fr

*E-mail addresses:* regis.lambert@upmc.fr (R.C. Lambert), Christine.malot@unice.fr (C. Tuleau-Malot), thomas.bessaih@upmc.fr (T. Bessaih), Vincent.Rivoirard@dauphine.fr (V. Rivoirard), Yann.bouret@unice.fr (Y. Bouret), nathalie.leresche@upmc.fr (N. Leresche), reynaudb@unice.fr (P. Reynaud-Bouret)

*Article Type:* Research article

## Abstract

**Background:** Statistical models that predict neuron spike occurrence from the earlier spiking activity of the whole recorded network are promising tools to reconstruct functional connectivity graphs. Some of the previously used methods are in the general statistical framework of the multivariate Hawkes processes. However, they usually require a huge amount of data, some prior knowledge about the recorded network, and/or may produce an increasing number of spikes along time during simulation.

**New Method:** Here, we present a method, based on least-square estimators and LASSO penalty criteria, for a particular class of Hawkes processes that can be used for simulation.

**Results:** Testing our method on small networks modeled with Leaky Integrate and Fire demonstrated that it efficiently detects both excitatory and inhibitory connections. The few errors that occasionally occur with complex networks including common inputs, weak and chained connections, can be discarded based on objective criteria.

**Comparison with existing methods:** With respect to other existing methods, the present one allows to reconstruct functional connectivity of small networks without prior knowledge of their properties or architecture, using an experimentally realistic amount of data.

**Conclusions:** The present method is robust, stable, and can be used on a personal computer as a routine procedure to infer connectivity graphs and generate simulation models from simultaneous spike train recordings.

(216 words)

*Keywords:* connectivity, spike train analysis, neuron correlation, lasso penalization, least-square estimation, Hawkes processes

# 1 Introduction

Ensemble spiking activities in neuron networks are the bases of information coding and information processing within the brain. Thanks to the significant advances in electrophysiological techniques ([5]) and computational power over the last 20 years, simultaneous spike train recordings have been collected in many brain areas under various experimental conditions from anesthetized (see for example [27]) to fully awake animals performing complex behavioral tasks (see for example [28]). Consequently, significant efforts are currently devoted to develop state of the art methods analyzing simultaneously recorded spike trains ([29, 26]). One of the main challenges faced by these methods is how to determine the potential interactions existing between the different neurons, *i.e.* to reconstruct the functional connectivity graph of the recorded network.

Commonly used approaches include several histogram-based methods to calculate the cross-correlation functions ([19]), the cross-intensity functions ([4]) or the joint peri-stimulus time histogram ([9]). These methods however only consider pairs of neurons and may easily lead to inaccurate functional connectivity graphs due, for example, to the presence of common inputs or connection chains within the recorded networks ([29]).

More recent methods are based on models of the whole set of observed spike trains ([29, 26]). Among these methods, much attention has been paid to statistical models that aim to predict spike occurrences in a given neuron as a function of its earlier spikes, of the preceding activities of the other recorded neurons, and possibly of some other known external variables, such as stimuli or movement ([4, 8, 14, 16, 17, 18, 20, 21, 22, 23]). Such models where mutual excitation, inhibition and synaptic integration play a key role are naturally described by the general probabilistic framework of the multivariate Hawkes processes. For such processes, the conditional intensity of point occurrence is a function  $\Phi$  of a linear combination of the background point rate and of multiple temporal kernels. These kernels usually materialize the potential impact of the preceding activities of every recorded neuron. The model is generally fitted using maximum likelihood estimation by determining the parameters of the kernels for which the actual recorded spike history has the highest probability. If nothing is known about the potential connectivity, the total number of parameters is typically of the order of  $pn^2$  where  $n$  is the number of spike trains and  $p$  the number of parameters per kernel. Therefore, if little is known in advance (shape of the kernels, connections to discard, etc.), the number of parameters to estimate is huge even for very small networks and several thousands of spikes per train are required to perform good estimations ([8]). Moreover the computational cost of maximum likelihood estimation can be very large even with only two recorded spike trains. Therefore, because of the shape of the log-likelihood for such processes, a popular choice is to use  $\Phi(\cdot) = \exp(\cdot)$ , which simplifies the computation (see for example [22]). In the neuroscience literature, such processes are often referred as (a particular case of) General Linear Models (GLMs). However, from a modeling point of view, it is possible that, once the parameters are estimated and plugged in simulation of the corresponding Hawkes processes, the amount of simulated spikes per unit of time increases with simulation time, a behavior that is not biologically relevant.

Our main aim is to propose a *turnkey* procedure, *i.e.* an efficient procedure that can be used routinely on laptops of experimenters without the help of any specialist, to reconstruct functional connectivity. Therefore, the procedure needs to be performed quickly at least when dealing with the relatively small number of simultaneously recorded spike trains that are usually analyzed in *in vivo* studies. Moreover, it has to be run without any prior knowledge or hypothesis regarding the recorded network, *i.e.* the few parameter settings that are left to the practitioner need to be perfectly understandable and calibrated in advance. Finally, the procedure should reconstruct a connectivity graph that comes with a fully estimated process, which can be simulated to reproduce realistic datasets.

In these respects, the recent method introduced in [24, 10], based on a least-squares criterion combined with a LASSO penalty, is very promising. The practitioner only needs to set one tuning parameter  $\gamma$  and to choose an appropriate bin size ( $\delta$ ) and number of bins ( $K$ ) (equivalently a range ( $K\delta$ )) to quantify the point process interactions. Here, using small Leaky Integrate and Fire networks, we evaluate for the first time if such Hawkes model approximations apply to neuronal spike trains and can efficiently reconstruct the underlying neuronal network functional connectivity. Challenging the method with complex network architectures, including chained connections and/or common inputs, drove us to propose an appropriate value for the tuning parameter  $\gamma$  and a correction of this LASSO method for Hawkes processes that makes it suitable to analyze realistic neuronal spike trains. The resulting procedure is effective with experimentally compatible amount of data, and is therefore proposed as a *turnkey* procedure to become a classical tool in the analysis of small numbers of simultaneously recorded neuronal activities.

The code required to run the present analysis method on simulated or recorded spike trains is implemented and parallelized in C++ and interfaced with R. It is available at <https://github.com/ybouretn/neuro-stat>.

## 2 Method

### Hawkes models

The  $n$  simultaneous spike trains, modeled as multivariate Hawkes processes, are usually characterized by their (conditional) intensity. More precisely, the intensity  $\lambda_i(t)$  of the  $i^{th}$  spike train  $N^i$  represents the probability to observe a new point around  $t$  for the spike train  $i$  given what already occurred on all the spike trains strictly before time  $t$ . Hence, the conditional intensity of a given point process can be seen as an instantaneous firing rate given the past events. For spike trains modeled by Hawkes processes, their intensities are of the following form

$$\lambda_i(t) = \left( \nu_i + \sum_{j=1}^n \sum_{T \in N^j, T < t} h_{j \rightarrow i}(t - T) \right)_+ ,$$

where  $(\cdot)_+$  is the positive part. The coefficient  $\nu_i$  is the spontaneous firing rate of the  $i^{th}$  spike train and the functions  $h_{j \rightarrow i}$  model the interaction of the  $j^{th}$  spike train on the  $i^{th}$  spike train. A large positive value of  $h_{j \rightarrow i}(d)$  at a certain delay  $d$  indicates an excitation from the neuron generating the  $j^{th}$  spike train onto the neuron generating the  $i^{th}$  spike train. As a consequence, when a spike occurs on the  $j^{th}$  spike train, the apparition of a new spike on the  $i^{th}$  spike train is very likely after a delay  $d$  (see Figure 1). Conversely, if  $h_{j \rightarrow i}(d) < 0$  with large value for  $|h_{j \rightarrow i}(d)|$ , the  $j^{th}$  spike train inhibits the apparition of new spikes on the  $i^{th}$  spike train after a delay  $d$ . By summing up all the possible interactions with all the possible delays, one obtain the current firing rate given the past events. It should be noted that the function  $h_{i \rightarrow i}$  represents the auto-interaction of the spike train and therefore models some intrinsic properties of the neuron such as the spike refractory period. Since the inhibition might be strong enough to generate negative  $\lambda_i$ , the overall positive part  $(\cdot)_+$  is considered to avoid negative firing rate. However in the sequel and to make the method intelligible, we will assume that the case of too strong inhibitions, given the past events, never occurs so that  $\lambda_i$  remains linear in both  $\nu_i$  and the  $h_{j \rightarrow i}$ 's, meaning that the positive part  $(\cdot)_+$  is useless.

The point measure  $dN_t^i$ , which counts 1 in a small interval if there is a spike for the  $i^{th}$  spike train in  $t$  and 0 otherwise, can therefore be seen as a noisy version of

$$\nu_i + \sum_{j=1}^n \sum_{T \in N^j, T < t} h_{j \rightarrow i}(t - T) \quad (1)$$

Therefore, modeling the  $n$  spike trains by Hawkes processes implies to estimate the parameters  $\nu_i$  and the functions  $h_{j \rightarrow i}$ . To parametrize this problem, we assume the functions  $h_{j \rightarrow i}$  to be piecewise constant on a partition of  $K$  bins of size  $\delta$  (Figure 1)

$$h_{j \rightarrow i} = \sum_{k=1}^K a_{j \rightarrow i}^k \mathbf{1}_{((k-1)\delta, k\delta]}$$

with  $a_{j \rightarrow i}^k$  and  $\nu_i$  in Hz. The spontaneous part  $\nu_i$  gives the average frequency of apparition of a new spike given that there is no past events. The coefficient  $a_{j \rightarrow i}^k$  gives the average frequency gain, or loss, of the  $i^{\text{th}}$  spike train induced by the occurrence of a spike on the  $j^{\text{th}}$  train in the preceding period between  $(k-1)\delta$  and  $k\delta$ .

Note that each term of the sum given by (1) can be rewritten as

$$\sum_{T \in N^j, T < t} h_{j \rightarrow i}(t - T) = \sum_{k=1}^K a_{j \rightarrow i}^k N_{[t-k\delta, t-(k-1)\delta]}^j$$

where  $N_A^j$  denotes the number of spikes in interval  $A$ . Therefore organizing the unknown parameters ( $\nu_i$  and  $a_{j \rightarrow i}^k$ ) in the vector  $\mathbf{a}_i$  of size  $1 + nK$ , one can rewrite (1) as

$$\nu_i + \sum_{j=1}^n \sum_{T \in N^j, T < t} h_{j \rightarrow i}(t - T) = \mathbf{c}_t^\dagger \mathbf{a}_i$$

where  $\dagger$  denotes the transposition and  $\mathbf{c}_t$  is an observable vector of size  $1 + nK$  that depends on  $t$ . In this scalar product, the first coordinate of  $\mathbf{c}_t$  is equal to 1 and is multiplied by the unknown  $\nu_i$ ; the other coordinates of  $\mathbf{c}_t$  are given by the number of past spike occurrences, so are in the form of  $N_{[t-k\delta, t-(k-1)\delta]}^j$  for some  $j$  and  $k$  and each of them is multiplied by the corresponding unknown  $a_{j \rightarrow i}^k$ .

### Least-squares estimation

The true vectors  $\mathbf{a}_i$  are unknown but since  $dN_t^i$  can be viewed as a noisy version of  $\mathbf{c}_t^\dagger \mathbf{a}_i$ , one can therefore minimize the distance between the  $dN_t^i$ 's and the  $\mathbf{c}_t^\dagger \beta$  over all the possible vectors  $\beta$  of size  $1 + nK$ . If the spike trains are observed between 0 and  $T_{max}$ , one wants therefore to minimize

$$\int_0^{T_{max}} (dN_t^i - \mathbf{c}_t^\dagger \beta dt)^2$$

Developing and keeping only the terms that depend on  $\beta$ , we end up with the following least-square criterion

$$LS^i(\beta) = -2 \int_0^{T_{max}} \mathbf{c}_t^\dagger \beta dN_t^i + \int_0^{T_{max}} \beta^\dagger \mathbf{c}_t \mathbf{c}_t^\dagger \beta dt = -2\mathbf{b}_i^\dagger \beta + \beta^\dagger \mathbf{G} \beta \quad (2)$$

where the matrix  $\mathbf{G}$  is defined by

$$\mathbf{G} = \int_0^{T_{max}} \mathbf{c}_t \mathbf{c}_t^\dagger dt,$$

and where  $\mathbf{b}_i$  is an observable vector of size  $1 + nK$ : its first coordinate is the number of spikes between 0 and  $T_{max}$  for the the  $i^{th}$  spike train and the other coordinates are given by

$$\int_0^{T_{max}} N_{[t-k\delta, t-(k-1)\delta]}^j dN_t^i,$$

for all  $j$  and  $k$ , that is in other words the cross-correlogram between the  $i^{th}$  spike train and the other  $j^{th}$  spike trains.

Therefore, minimizing in  $\beta$ , the least-squares estimate of the parameter  $\mathbf{a}_i$  is given by

$$\hat{\mathbf{a}}_i = \mathbf{G}^{-1} \mathbf{b}_i \quad (3)$$

Note that the classical direct estimation of interactions between  $j$  and  $i$  that is performed via cross-correlograms appears in  $\mathbf{b}_i$ . However, in many cases the simple cross-correlogram analysis can be misleading. For instance, a strong chain of excitation between 3 neurons as  $1 \rightarrow 2 \rightarrow 3$  might result in cross-correlograms that suggest a non-existing direct excitation from 1 to 3. In equation (3), the matrix  $\mathbf{G}$  takes into account such potential complex relationships between the different recorded spike trains and its inversion untangles these links (see also the numerical experiment in [24]) for Hawkes processes. However, as it is, the method is not optimal for two reasons:

- the number of parameters to be estimated is huge. For example, simultaneous recordings of  $n = 10$  spike trains, considering  $h_{j \rightarrow i}$  functions over a range of 50ms divided in  $K = 10$  bins, require the estimation of 1010 parameters. Therefore, the observation duration  $T_{max}$  is usually not long enough to calculate a good estimation of such a huge number of parameters.
- many parameters are potentially irrelevant, in particular many  $h_{j \rightarrow i}$  functions are generally null. Therefore, one wants to estimate a sparse graph of interactions revealing the connections that really exist.

## LASSO estimation

To stabilize the method as well as gain sparsity, we use a two-step LASSO method and first minimize the following criterion

$$LASSO^i(\beta) = -2\mathbf{b}_i^\dagger \beta + \beta^\dagger \mathbf{G} \beta + 2\mathbf{d}_i^\dagger |\beta|, \quad (4)$$

where  $|\beta|$  is the vector whose coordinates are the absolute values of  $\beta$  and where  $\mathbf{d}_i$  is a vector of weights. The theory presented in [10] shows in a much more general framework that multivariate counting processes are highly heteroscedastic (i.e. the variance depends on the coefficients) and that therefore the weights need to be set accordingly to obtain a robust method. In particular [10] proposes data-dependent weights that are based on an estimation of the variance on the data themselves. By simplifying the method fully detailed in [10], we end up with (the following equality should be understood coordinate by coordinate)

$$\mathbf{d}_i = \sqrt{2\gamma \log(n + n^2 K) \int_0^{T_{max}} \mathbf{c}_t^2 dN_t^i} + \frac{\gamma \log(n + n^2 K)}{3} \sup_{t \in [0, T_{max}]} |\mathbf{c}_t|.$$

Note that these weights, and therefore the presented LASSO method itself, depend on only one tuning parameter  $\gamma$ . For a fixed value of  $\gamma$ , the minimization of  $LASSO^i(\beta)$  leads to an estimate  $\hat{\mathbf{a}}_i^{BL}$  of the true parameters  $\mathbf{a}_i$ . "BL" stands for "Bernstein Lasso" since the shape of the weights is derived by a Bernstein concentration inequality.

The derivation of the weights in [10] shows that  $\gamma$  : (i) qualitatively controls the probability that the estimator is good, (ii) should be an absolute constant that does not depend on the problem at hand and (iii) can be tuned in advance. Note that  $\gamma = 0$  corresponds to the least-square estimate  $\hat{\mathbf{a}}_i$  which is not sparse at all and that the sparsity (that is the number of null coefficients) increases with  $\gamma$ . As a consequence, reconstructed networks associated with  $\gamma$  too small will contain non-existing interactions, whereas taking  $\gamma$  too large will not detect all existing interactions. Therefore, the first required step to use the present method in spike train analysis is the choice of a reasonable value for  $\gamma$  (see Results).

Since the first applications of LASSO methods [30], it is well known that, if a true sparse set of parameters underlying the observed data exists, the LASSO estimator, once correctly tuned, will find the set of non-zero parameters if enough observations are provided. In our case, it means that if the spike train dataset indeed obeys a Hawkes process with a sparse vector  $\mathbf{a}_i$ , then  $\hat{\mathbf{a}}_i^{BL}$  will be null exactly where the true parameter vector is. In particular, this implies that if the data are generated by such a Hawkes process, then the reconstructed connectivity graph which is obtained by keeping an edge from the  $j^{th}$  to the  $i^{th}$  spike train as soon as there is one non-zero  $|\hat{\mathbf{a}}_{j \rightarrow i}^{BL,k}|$  is the correct one. But it is also well known [30] that the LASSO reconstruction of the non-zero coefficients is not optimal. This can be easily corrected by performing a second step of parameter estimation, where the Ordinary Least Square estimator is restricted to the coordinates where the  $|\hat{\mathbf{a}}_{j \rightarrow i}^{BL,k}|$ 's are non zero. Since the LASSO properties guarantee that there is a very small number of such coefficients, the variance of such a least-square is much smaller than the full least-square estimator. This second estimator is denoted  $\hat{\mathbf{a}}_i^{BOL}$ .

Using  $\hat{\mathbf{a}}^{BOL}$ , it is therefore possible to estimate the interaction strength of a given connection  $j \rightarrow i$  in the connectivity graph, which is defined by

$$\int h_{j \rightarrow i}(x) dx = \sum_{k=1}^K a_{j \rightarrow i}^k \delta$$

This interaction strength gives the overall average number of spikes gained (if positive) or lost (if negative) on the  $i^{th}$  train following the occurrence of a spike on the  $j^{th}$  train at any delay. However, since the occurrence of spikes in the  $j^{th}$  train may have complex effects combining both excitation and inhibition at different delays on the  $i^{th}$  spike train, the net modification in the number of spikes in the  $i^{th}$  train can be close to 0 in cases of strong combined excitatory and inhibitory interactions.

Therefore we also introduce the overall *energy of an interaction* as the average number of spikes of the  $i^{th}$  spike train that has been modified (either *gain* or *loss*) by the presence of a spike on the  $j^{th}$  train at any delay, which is mathematically defined by

$$\int |h_{j \rightarrow i}(x)| dx = \sum_{k=1}^K |a_{j \rightarrow i}^k| \delta$$

### 3 Results

If  $\gamma$  is fixed and large enough, the method was shown to achieve very good performance in recovering sparse vectors of parameters when point processes are generated by Hawkes processes (see [10]). We refer the interested reader to the simulation study in [10] for a comparison with other LASSO methods. However, the method has never been tested with Hawkes processes that simulate classical difficulties encountered in neuronal spike train analysis. Therefore, as a first step to evaluate its efficiency and provide an estimation of the  $\gamma$  value, we analyzed data obtained from simulations of small networks in which each neuron was modeled as a true Hawkes process. Note that, since our



aim is to provide a method that is not too computationally intensive and can already reconstruct connectivity graphs with a small number of datasets, we focus here on determining a reasonable tuning value for  $\gamma$  instead of developing an automatic selection procedure as cross-validation methods.

**Calibration of  $\gamma$  using Hawkes process simulation.** We used a 4 neuron network where 2 neurons received a strong excitatory input from a third one (see *Connectivity graph* in Figure 2A). This generates a classical difficulty in spike train analysis since the spike correlation of the 2 neurons receiving the strong common input may create false non-existing links in the reconstructed connectivity graph. More precisely, to design a Hawkes network model that behaves like a classical neuronal network, we first constructed a template network with 4 Leaky Integrate and Fire (LIF). Each LIF presented a spontaneous activity set to 11Hz and two of them shared a common input with a weight equal to 0.7 (see below for details on LIF). The parameters of the best Hawkes model approximation of this LIF network were then estimated by  $\hat{\mathbf{a}}^{BOL}$  and used to construct the Hawkes model corresponding to the connectivity graph of Figure 2A.

In this framework, where only Hawkes processes are simulated, we first investigated the impact of different  $\gamma$  values on the connectivity graph reconstructions. Note that perfect reconstruction of such Hawkes process network could be especially challenging due to the overall strength of the interaction functions that is larger than 1. Indeed, more than 1 spike is created on average in the target neurons when the input neuron fires. For each  $\gamma$ , the reconstruction efficiency was evaluated as a function of the dataset duration by calculating in each case the number of graphs that were perfectly reconstructed at the end of 100 dataset analysis. As shown in Figure 2A, the LASSO method (with  $\gamma$  around 3 or larger) recovers the right connectivity graph in a large majority of simulations as soon as the dataset duration is larger than 10 to 20s.

In these analyses, the  $h_{j \rightarrow i}$  functions were evaluated with  $K = 10$  bins of  $\delta = 5$ ms. A closer look at the shape of the interaction functions (Figure 2B) indicates that, as expected, the estimation of the interaction functions improves with the dataset duration. However, modifying  $K$ ,  $\delta$  and the overall support  $K\delta$  does not change the overall shape of the interaction functions nor the connectivity graph. Indeed, the sparsity of  $\hat{\mathbf{a}}^{BOL}$  implies a sparsity of the connectivity graph but also a in the range of the interaction functions. Therefore if a parameter  $K$  larger than required to accurately describe the interaction function is provided, the non-necessary estimated  $a_{j \rightarrow i}^k$  are set to zero and the effective range remains qualitatively unchanged.

**Hawkes-model based detection of excitatory and inhibitory connections within LIF networks.** It has been shown in [10] that, when the underlying processes are not Hawkes processes, like in case of real neuronal spike trains, the present method fits the "best Hawkes approximation", which is sparse. In this case, the estimated intensity functions should rightly approximate the probability to see a new point around the time  $t$  but the connectivity graph may be wrong. Therefore the robustness of the present method in terms of connectivity reconstruction has to be further tested using classical model networks of known architectures that mimic neuronal spike generation and are not Hawkes models.

Hence, to evaluate the efficiency of the present method in detecting spike train interactions, artificial networks of 10 spontaneously active LIF were constructed using the IntFire4 class of the Neuron 7.4 environment ([13]). Briefly, artificial neurons integrated fast monoexponentially decaying ( $\tau_e = 3$ ms) excitatory or slower alpha function like biexponential ( $\tau_{i1} = 5$ ms,  $\tau_{i2} = 10$ ms) inhibitory inputs with a membrane time constant  $\tau_m$  (30ms) and fired when the membrane state variable reached the threshold 1. After firing, the membrane state was set to 0. Excitatory events were scaled such that an isolated event of weight 1 produced a maximum membrane potential of 1

(threshold) and an isolated inhibitory event of weight -1 produced a minimum membrane potential of -1 (see details in [6]). Each LIF was driven by a Poisson excitatory input to induce spontaneous firing activity of 11Hz.

The first simulated networks only contained a single excitatory or inhibitory connection inserted between 2 of the neurons (Figure 3A) and simulations of various duration (from 20s to 360s) were run with increasing connection weights (from 0.1 to 1 and -0.4 to -1.4 for excitatory and inhibitory connections, respectively). In each condition, the analysis was performed on 100 datasets in order to assess the reproducibility of the results.  $\gamma$  was systematically increased from 0.02 to 6 (on a non-regular grid) in order to determine whether a value for  $\gamma$  can be fixed that allows to rightly detect excitatory or inhibitory connections between LIF neurons. As presented in the *Method*, if  $\gamma$  is too small, the estimation procedure should provide reconstructions based on Hawkes models with non-null interaction functions that do not correspond to existing connections in the neuronal network. We, therefore, calculated in each condition the False Positive Rate (FPR) as the percentage of datasets where an additional interaction (i.e. a non zero interaction function  $h_{j \rightarrow i}$  that does not correspond to an existing interaction between the simulated neurons) was found. As presented in the Figure 3B showing for each condition the smallest value of  $\gamma$  (noted  $\gamma_{lim}$ ) that guarantees an FPR inferior to 5%, values larger than 2.15 for excitatory interactions and 2.45 for inhibitory interactions should be selected in order to avoid reconstructions suggesting non-existing connections. Note that our aim is to determine a fixed  $\gamma$  value that guarantees good performances (at least in terms of FPR) whatever the model, the interaction strength or the observation duration. Therefore, considering  $\gamma$  equal or larger than 3, as suggested by the previous analysis of Hawkes model networks, should be an accurate choice when using the present method to analyze more realistic spike trains.

However, since larger values for  $\gamma$  may favor Hawkes models with "missing connections" (i.e. non-detected connections), we calculated in a second step the True Detection Rate (TDR) as the percentage of datasets where the functional connectivity graph was perfectly estimated for  $\gamma = 3, 3.5$  and 4 (Figure 3C). For both excitatory and inhibitory connections, the TDR curves are similar for the three  $\gamma$  values, with a slightly better performance for  $\gamma = 3$ . As shown in Figure 3C1, although the TDR is larger for stronger excitatory connections since a high connection weight increases the proportion of spikes due to the excitatory inputs in the analyzed data, detection of excitatory interaction appears highly efficient in every condition. Hence, with approximately 1350 spikes spontaneously generated in each LIF during 120s long datasets, network reconstructions were already perfect in every trial with a connection weight of 0.15 that only added about 230 supplementary spikes to the target LIF, corresponding to a 17% increase in firing rate. However, as for any statistical methods, detection efficiency drops dramatically with shorter datasets such that, for 20s long duration where about 230 spontaneously generated spikes / LIF were available to perform the analysis, the excitatory connection had to increase the firing frequency by at least 63% for accurate detection in every trial.

For obvious statistical reasons, inhibitory inputs are trickier to detect than excitatory ones. Indeed, in the analyses based on extracellular recordings, the presence of an inhibitory connection is revealed by missing spikes, *i.e.* spikes that are suppressed due to the inhibitory input. Such spike suppression may be rare if the intrinsic firing rate of the target neuron is low or the inhibitory input weak. Therefore, accurate detection of inhibitory inputs requires a large amount of data and a sufficient probability of spike suppression. Accordingly, although perfect detection was already achieved in every trial with a dataset of 160s (about 1800 spontaneously generated spikes / LIF), a strong inhibitory connection weight of -0.6 is required to decrease the firing frequency of the target neurons by 61% (Figure 3C2). An additional difficulty to detect inhibitory connection is illustrated by the shapes of the TDR curves that are not monotone functions of the inhibitory input weight. Indeed very strong inhibition can paradoxically make the target neuron almost silent and therefore invisible

in the dataset, inducing a drastic decrease in the detection efficiency.

Nevertheless, we conclude from these first series of simulations that setting  $\gamma = 3$  is a compromise that minimizes the probability to reconstruct networks with non-existing and/or missing connection from spike trains generated by either Hawkes (see above) or LIF models.

In the previous analysis, the  $h_{j \rightarrow i}$  functions were evaluated with  $K = 10$  bins of  $\delta = 5\text{ms}$ . As long as the choice of the support  $K\delta$  remains reasonable with respect to the kinetics of the excitatory or inhibitory events, modifying the  $K$  or  $\delta$  values does not change the reconstructed connectivity graph for a given value of  $\gamma$ . However, as shown in Figure 4, if one is interested in the precise shape of the  $h_{j \rightarrow i}$  functions the length of the support should be chosen to largely exceed the maximal duration of the synaptic events. Indeed, with our initial choice of  $K = 10$  bins of  $\delta = 5\text{ms}$  (support of  $50\text{ms}$ ), the Lasso method accurately estimates the fast excitatory inputs setting the nonrelevant  $a_{j \rightarrow i}^k$  coefficients to zero. Decreasing  $\delta$  while keeping a large enough support does not drastically change the  $h_{j \rightarrow i}$  function (Figure 4A). However, since the inhibitory events have a much slower kinetics (biexponential decay with time constants,  $\tau_{i1} = 5\text{ms}$  and  $\tau_{i2} = 10\text{ms}$ ) than the excitatory ones (monoexponential decay with time constant,  $\tau_e = 3\text{ms}$ ), our initial support of  $50\text{ms}$  was not long enough to properly describe the corresponding  $h_{j \rightarrow i}$  function. This function was therefore truncated (Figure 4B). Nevertheless, note that such truncation did not impede the reconstruction of the connectivity graph (see blue line in the *left panel* of Figure 3C2).

**”Energy” of the  $h_{j \rightarrow i}$  functions helps to discard non-existing connections.** The previous simulations show that the LASSO method based on Hawkes models correctly infers excitatory relationships from spike trains within simple small LIF networks. In a second step, we investigated the success rate of the method when it was challenged with classical difficulties in spike correlation analysis such as large differences in firing frequencies or the presence of common inputs in the network. The values  $K = 10$  bins of  $\delta = 5\text{ms}$  were used in order to keep conditions that provide good detection of excitatory connections.

So far, the intrinsic firing rates of every LIF in our simulated networks were similar and fixed around  $11\text{Hz}$ . However, two independent neurons recorded simultaneously and displaying very different firing rates may present apparent spike correlations if, due to its high firing probability, the fastest neuron often fires at the same time than the slowest one. In such cases, the most favorable Hawkes model that fits the data may include a false excitatory connection linking the two neurons. Therefore, simulating non-connected LIFs among which one neuron presented an up to 20 times faster firing rate than the other ones, we evaluated in each case  $\gamma_{lim}$ , the minimal value for  $\gamma$ , that guaranteed an FPR inferior to 5%. As shown in Figure 5A,  $\gamma = 3$  could not assure an FPR inferior to 5% in every case, in particular for long datasets that contained many spike coincidences. From a mathematical point of view, it may seem counter-intuitive that error occurrences increased with the dataset duration. However, LIF and neuron firings are not true Hawkes processes and reducing the data variance effect with long datasets makes the bias of the model more apparent. Nevertheless, the method was surprisingly robust since for every tested duration, FPR inferior to 5% was still achieved with  $\gamma \leq 3$  for simulations where one of the LIFs was firing almost ten times faster than the other ones.

Another classical difficulty in spike train analysis arises when recorded neurons are serially connected with strong excitatory links since correlation may exist between the spikes of the first and last neurons of the chain although no direct connection is present. A similar problem is met when neurons share a strong common excitatory input that creates spike correlation in the target neurons although these are not directly connected. Therefore we systematically investigated, as a function of the dataset duration and connection weight, the ability of our model-based method to infer connection graphs in networks comprising serial connections (Figure 5B) or shared inputs (Figure

5C). In both cases, no value for  $\gamma$  can be found on the grid that provides Hawkes reconstruction with false connection in less than 5% of the analysis as soon as the weight of the connection is very large. If we consider in more details the network of Figure 5C with  $w = 0.7$  and datasets of 360s, In almost every simulation, although the strength of the corresponding reconstructed interaction functions was very small, additional edges between LIF 2 and 4 were present in the approximated Hawkes model even for  $\gamma = 6.5$ . However, we demonstrated above (see *Calibration of  $\gamma$  using Hawkes process simulation*) that when simulating the same network with Hawkes processes, the present method perfectly reconstructs the connectivity graph without additional edges with  $\gamma = 3$  (Figure 2). Our interpretation is that, in case of data that are not generated by true Hawkes processes, the LASSO method indeed fits the best Hawkes approximation of the intensity when the observation time is long enough (as mathematically proved in [10]) but the reconstructed connectivity graph of this approximation does not always coincide with the true connectivity graph. When the observation time is long enough, all the statistical fluctuations have vanished, being completely controlled by the LASSO penalty with  $\gamma = 3$ , but the bias of approximating the data by a Hawkes model remains. This is precisely the phenomenon that needs to be anticipated to make the method robust when real data are analyzed.

To increase the robustness of our procedure, we further analyzed the functions  $h_{j \rightarrow i}$ . As already stated, it clearly appears that, when reported, the additional interactions display a very low "energy" compared to existing connections (Figure 6). Therefore, systematically discarding connections with "energy" smaller than an arbitrary threshold allows to almost perfectly suppress the additional false connections of the LASSO method with  $\gamma = 3$  (Figure 5B and C, middle graphs). However, although discarding low "energy" functions always improves reconstruction of the true connection graph, optimizing the choice of the discarded connection requires to adapt the threshold value in each case. Different rules can be designed to fix such a threshold value and we report here a correction of the LASSO method that produced good results in the relatively simple simulations used in the present paper.

This correction is based on the strength of the interaction functions rather than on their energy since we never observed in our simulations any additional inhibitory connections (as illustrated in the next paragraph). Therefore, we chose a correction parameter that easily distinguishes excitatory from inhibitory connections. More precisely, the values  $E_{j \rightarrow i} = \int h_{j \rightarrow i}$ , which represent the number of additional points created in average in the spike train  $i$  thanks to the presence of one spike on the  $j^{th}$  spike train, were computed from the coefficients  $\hat{\mathbf{a}}^{BOL}$  and sorted in increasing order. The first large "jump" in the ordered sequence of  $E_{j \rightarrow i}$  can then be used to set the threshold value. In practice, the successive differences in the sorted sequence of  $E_{j \rightarrow i}$  are calculated and the first difference superior to 15% of the largest one designs the first large "jump". Accordingly, the excitatory  $h_{j \rightarrow i}$  functions corresponding to the values  $E_{j \rightarrow i}$  smallest than this first large "jump" are discarded (Figure 7). When applied to the estimation in the previous simulations (Figure 5B and C, bottom graphs), this first large "jump" correction allowed us to perfectly reconstruct the simulated network in every case.

**Functional connectivity graphs.** The ability of the present method (with  $\gamma=3$  ;  $K=20$  and  $\delta=5\text{ms}$  in order to fix an appropriate support for both excitatory and inhibitory interaction functions) to reconstruct functional connectivity graphs was finally tested using randomly generated small networks of LIFs ( $n = 7$ ). Each LIF presented a spontaneous firing rate selected at random in  $[10, 40]$  Hz. Connection from the  $j^{th}$  to the  $i^{th}$  LIF had a  $1/n$  probability in order to ensure a sparse graph according to Erdős-Renyi random graph properties. Each existing connection was determined as excitatory or inhibitory with a probability taken at random on  $[0, 1]$  and fixed for each graph. Excitatory and inhibitory weights  $w_{j \rightarrow i}$  were set uniformly at random in  $[0.2, 0.5]$  and  $[-0.9, -0.6]$ ,

respectively. Among the 100 randomly designed networks that were simulated, 11 networks generated less than 8 spikes in 7 minutes due to strong global inhibition and were thereafter discarded. We then counted the number of errors, either additional or missing excitatory/inhibitory connections, in the 89 remaining functional connectivity graphs that were reconstructed (Figure 8A). Non-existing inhibition was never observed and increasing the dataset duration improved the detection of inhibitory connections, as already mentioned. As a consequence, the proportion of networks in which inhibition was perfectly inferred drastically increased from 30 to 75% with the dataset duration (Figure 8B, left graph). However, although long dataset duration similarly improved the detection of existing excitatory connections, it also favored Hawkes approximations with non-existing excitatory connections (Figure 8A). Therefore, the improved detection of existing excitatory connections being counterbalanced by the appearance of non-existing ones (see green line in the left graph of Figure 8B), the proportion of networks with perfectly detected excitation remained almost stable above 60%. Hence, refining the functional excitatory connectivity graphs ( $h_{j \rightarrow i}$  functions with  $E_{j \rightarrow i} = \int h_{j \rightarrow i} > 0$ ) by discarding functions using the first large "jump" correction significantly increases the proportion of perfectly detected excitatory networks up to 80% (Figure 8B, right graph).

**Computational time.** As shown in Figure 9, the analysis presented for about ten spike trains and  $K = 10$  can be run effectively in a few seconds on a classical laptop. Note however that the complexity of the algorithm is mainly driven by the size of the matrix  $\mathbf{G}$ , *i.e.*  $K^2 n^4$ . Since the number of threads on a classical laptop is usually small, any multiplication by 10 of the number of neurons would multiply the computational time by about 10000. Hence the method can indeed be used as a routine procedure on laptops as long as the number of spike trains remains reasonable.

## 4 Discussion

We will briefly discuss below the rationale that presides to the design of the present method modeling spike trains with multivariate Hawkes processes.

The general expression of the intensity of Hawkes processes is given by

$$\lambda_i(t) = \Phi \left( \nu_i + \sum_{j=1}^n \sum_{T \in N^j, T < t} h_{j \rightarrow i}(t - T) \right),$$

where  $\Phi$  is a fixed function. For linear Hawkes processes initially introduced by Hawkes in the 70's to model earthquakes and their aftershocks [11],  $\Phi$  is the identity ( $\Phi(t) = t$  for any  $t$ ) but from a probabilistic point of view, such processes can only be used when every interaction function  $h_{j \rightarrow i}$  corresponds to excitation to preclude the possibility of negative intensity. To solve this problem when modeling networks with both excitatory and inhibitory connections, we considered the positive part  $\Phi(\cdot) = (\cdot)_+$  instead of the identity.

One of the main desirable properties of such processes is the stationarity. When a point process is stationary, the distribution of the points in a neighborhood of  $t$  does not depend on  $t$ , whatever the  $t$ . Moreover, the stationary distribution is typically achieved by letting the system evolve for a long time. In this case, even if the observed biological system is not stationary *per se*, the model of this system somehow goes back to a stable behavior, the stationary regime, after a reasonable delay following any perturbation. One can prove that linear Hawkes processes can be in a stationary regime if and only if the largest eigenvalue of the matrix ( $\int h_{j \rightarrow i}$ ) is smaller than 1 [12]. If this

condition is not fulfilled, these processes generate an exponentially increasing number of spikes when the simulation time increases, which obviously precludes the use of such Hawkes processes to simulate neuronal activity. For more general Hawkes processes, Brémaud and Massoulié ([2]) showed that if  $\Phi$  is  $k$ -Lipschitz (i.e. the slope of every segment linking 2 points of the graph of the function  $\Phi$  is less than  $k$ ) with  $k \leq 1$ , and if the largest eigenvalue of the energy matrix ( $\int |h_{j \rightarrow i}|$ ) is strictly smaller than 1, then such Hawkes processes can be in a stationary regime. These conditions are easily fulfilled in our model using  $\Phi(\cdot) = (\cdot)_+$ . This ensures that our method generates stationary models that can be used not only to reconstruct the functional connectivity graph from spike trains but also to simulate datasets without producing an exponentially growing number of spikes during simulation.

In counting processes, parameter estimation is generally performed using methods based on maximum likelihood since, under some regularity conditions, maximum likelihood estimators produce the smallest asymptotic variance [31]. However, they are generally difficult to compute. Therefore to efficiently optimize linear Hawkes models with experimentally manageable datasets, we rather used least-square estimators that are consistent and much easier to derive than maximum likelihood estimators. Due to the presence of the log function in the maximum likelihood equation, some authors made a different choice and still used Hawkes models with maximum likelihood estimators and  $\Phi(\cdot) = \exp(\cdot)$  (see [22, 18, 7]) to facilitate parameters estimation. However, as explained above, such Hawkes models do not fulfill the stationary conditions identified so far and therefore may produce an exponentially growing number of spikes during simulations. Despite the computational difficulty, maximum likelihood estimators of linear Hawkes processes have nevertheless been previously used on spike trains [8]. However, good performances were only obtained with more than thousands of points per spike train precluding the use of this approach to analyze a majority of experimental data. An alternative method especially designed to optimize linear Hawkes process relies on the inversion of an empirical version of the Wiener Hopf system that couples the matrix of interaction functions to the infinitesimal covariance matrix ([1]). Although such methods also require a huge dataset to make the empirical version of the Wiener Hopf system close enough to the expected one, similar approaches have been previously used in neurosciences ([20, 21, 14, 16]).

Due to experimental constraints, the number of unknown parameters in the model is large with respect to the size of the dataset and the output variability of the optimization procedure is high. To reduce this variability, we apply a penalized criterion by looking for a sparse parameter vector, assuming that many of the unknown coefficients should be null. In the classical " $\ell_0$  penalty" approach, the final contrast (either the log-likelihood or the least square contrast) of a given parameter vector is penalized by a function of the number of non zero coefficients. It is commonly accepted that this method efficiently approximates the right non zero coefficients as long as the number of parameters is relatively small and fixed with respect to the number of observations. In [25], we demonstrated that up to a small modification of the " $\ell_0$  penalty", this penalization coupled with the least-squares contrast can also be used for a large number of parameters and applied even if this number is large with respect to the size of the available data. However, the computation cost of the minimization of this non-convex criterion is prohibitive. To circumvent this problem, we proposed to use an " $\ell_1$  penalty", namely, we penalize the contrast by the  $\ell_1$ -norm of the coefficients rather than by the number of non zero coefficients (see also [22] for other  $\ell_1$ -type penalties). This convex criterion, known as a LASSO criterion, was first introduced by Tibshirani in regression [30], and we previously proved ([10]) that one should use a weighted " $\ell_1$  penalty", like the one described in (4), to derive a calibrated procedure that adapts to the problem heteroscedasticity, *i.e.* the potential large differences in the variance of the estimated parameters.

It is worth mentioning that few other models, close to Hawkes processes, were previously used to assess functional connectivity. One of the initial work was performed by Brillinger and coauthors

([3, 4]) who successfully identified interactions between pairs and triplets of neurons in nonrecurrent Aplysia networks. However, it appears difficult to apply this method to multiple simultaneously recorded spike trains without prior knowledge of the network connectivity. More recently, Cox models with classical Cox estimations, pseudolikelihood, and corresponding tests were used to assess the existence of neuron connectivity ([17]). Such model may be considered as Hawkes models (with  $\Phi(\cdot) = \exp(\cdot)$ ) where the influence of only the first preceding spikes, and not the whole history of the neuron spiking activities, is considered to determine the intensity functions. Furthermore, the optimization procedure implied to limit the number of estimated parameters, a strong parametric assumption that may involve prior knowledge of the studied neuron networks. Pouzat and Chaffiol ([23]) developed similar approaches (Wold model) considering the distances to the two preceding spikes, and estimated the "interaction" functions using log-likelihood criterion penalized by Wahba regularization. However, the method is computationally intensive and would be difficult to apply to more than 2 simultaneously recorded spike trains on classical laptops.

The results of [10] show that the LASSO method should provide the best sparse Hawkes approximation of the intensity once the fixed parameter  $\gamma$  is large enough and this even if the underlying process is not a Hawkes process. However, we showed here that the corresponding reconstructed connectivity graph does not match, in some difficult but classical cases, the real connectivity graph when the data are not simulated according to a Hawkes process. This caveat is not due to a high level of noise since it persists and is even more present with long observation duration. Our simulation study shows that it is consistent with a bias due to the discrepancy between the model underlying the data and a true Hawkes process. This phenomenon, revealed here with LIF networks, will obviously be present when analyzing real neuronal data but we demonstrated that correction based on the strength of the interactions will efficiently correct the obtained graph.

## 5 Conclusion

In conclusion, when challenging LIF models of small neuron networks, the present method based on multivariate Hawkes processes and the optimization of least squares criteria combined with a LASSO penalization and a correction method based on the integral of the functions  $h_{j \rightarrow i}$ , efficiently retrieves the hidden functional connectivity. The method is robust, stable, does not require huge amount of data nor additional parameter constraints and can be quickly run on a personal computer. Therefore, we propose that it can be used as a routine *turnkey* procedure to infer potential connectivity graphs and to generate simulation models from simultaneous spike train recordings.

**Acknowledgments :** This work was supported by the French government, through the UCA<sup>JEDI</sup> "Investissements d'Avenir" managed by the National Research Agency (ANR-15-IDEX-01), the grant ANR 2011 BS01 010 01, the CNRS funding PEPS BMI 2012-2013, and by the interdisciplinary axis MTC-NSC of the University of Nice Sophia-Antipolis.

## References

- [1] Bacry, E., Jaisson, T. and Muzy, J.-F. *Estimation of slowly decreasing Hawkes kernels: Application to high frequency order book modelling*. Quantitative Finance, **16**(8), 1179–1201 (2016).
- [2] Brémaud, P. and Massoulié, L. *Stability of nonlinear Hawkes processes*. Ann. Prob. **24**(3), 1563–1588 (1996).
- [3] Brillinger, D. *The identification of point process systems*. Annals of Probability, **3**(6), 909–924 (1975).
- [4] Brillinger, D., Bryant, H. and Segundo, J. *Identification of synaptic interactions*. Biol. Cybernetics **22**, 213–228 (1976).
- [5] Buzsáki, G. *Large-scale recording of neuronal ensembles* Nat Neurosci, **7**, 446–451 (2004).
- [6] Carnevale, N.T. and Hines, M.L. *The NEURON book*. Cambridge: Cambridge University Press (2006).
- [7] Carstensen, L., Sandelin, A., Winther, O. and Hansen, N.R. *Multivariate Hawkes process models of the occurrence of regulatory elements*. BMC Bioinformatics (2010).
- [8] Chornoboy, E., Schramm, L. and Karr, A. *Maximum likelihood identification of Neural Point Process Systems*. Biol. Cybernetics, **59**, 265–275 (1988).
- [9] Gerstein, G.L. and Perkel, D.H. *Simultaneously recorded trains of action potentials: analysis and functional interpretation*. Science, **164**, 828–830 (1969).
- [10] Hansen, N.R., Reynaud-Bouret, P. and Rivoirard, V. *Lasso and probabilistic inequalities for multivariate point processes*. Bernoulli, **21**(1), 83–143 (2015).
- [11] Hawkes, A. G. Point spectra of some mutually exciting point processes. Journal of the Royal Statistical Society. Series B. Methodological, **33**, 438–443 (1971).
- [12] Hawkes, A. G. and Oakes, D. A cluster process representation of a self-exciting process. Journal of Applied Probability, **11** (3), 493–503 (1974).
- [13] Hines, M.L. and Carnevale, N.T. *The NEURON simulation environment* Neural Comput., **9**, 1179–1209 (1997).
- [14] Jovanović, S. and Rotter, S. *Interplay between Graph Topology and Correlations of Third Order in Spiking Neuronal Networks*. PLoS Computational Biology (2016).
- [15] Karabash, D. *On stability of Hawkes process*, arXiv:1201.1573v4.
- [16] Krumin, M., Reutsky, I. and Shoham, S. *Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input*. Frontiers in Computational Neuroscience, **4**, article 147 (2010).
- [17] Masud, M.S. and Borisyuk R. *Statistical technique for analyzing functional connectivity of multiple spike trains* Journal of Neuroscience Methods **196**, 201–219 (2011).
- [18] Okatan, M., Wilson, M. A. and Brown, E.M. *Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity*, Neural Computation, **17**, 1927–1961 (2005).



- [19] Perkel, D.H., Gerstein, G.L. and Moore, G.P. *Neuronal spike trains and stochastic point processes II. Simultaneous spike trains*. Biophys J., **7**, 419–440 (1967).
- [20] Pernice, V., Staude, B., Cardanobile, S. and Rotter, S. *How structure determines correlations in neuronal networks*. PLoS Computational Biology, 7:e1002059 (2011).
- [21] Pernice, V., Staude, B., Cardanobile, S. and Rotter, S. *Recurrent interactions in spiking networks with arbitrary topology*. Physical review E, Statistical, nonlinear, and soft matter physics, 85:031916 (2012).
- [22] Pillow, J., Shlens, J., Paninski, L., Sher, A., Chichilnisky, E. and Simoncelli, E. *Spatio-temporal correlations and visual signalling in a complete neuronal population*, Nature, **454**, 995–999, (2008).
- [23] Pouzat, C. and Chaffiol, A., *Automatic Spike Train Analysis and Report Generation. An Implementation with R, R2HTML and STAR*, Journal of Neuroscience Methods, **181**, 119–144 (2009).
- [24] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. *Inference of functional connectivity in Neurosciences via Hawkes processes*, 1st IEEE Global Conference on Signal and Information Processing, Austin, Texas, (2013).
- [25] Reynaud-Bouret, P. and Schbath, S. *Adaptive estimation for Hawkes processes; application to genome analysis*. Ann. Statist., **38**(5), 2781–2822 (2010).
- [26] Roudi, Y., Dunn, B. and Hertz, J. *Multi-neuronal activity and functional connectivity in cell assemblies*. Current Opinion in Neurobiology, **32**, 38–44 (2015).
- [27] Roy, N.C., Bessaih, T. and Contreras, D. *Comprehensive mapping of whisker-evoked responses reveals broad, sharply tuned thalamocortical input to layer 4 of barrel cortex*. J Neurophysiol., **105**(5), 2421–2437 (2011).
- [28] Schmitt, L.I., Wimmer, R.D., Nakajima, M., Happ, M., Mofakham, S. and Halassa, M.M. *Thalamic amplification of cortical connectivity sustains attentional control*. Nature, **545**(7653), 219–223 (2017).
- [29] Stevenson, I.H., Rebecco, J.M., Miller, L.E. and Krding, K.P. *Inferring functional connections between neurons*. Current Opinion in Neurobiology, **18**(6), 582–588 (2008).
- [30] Tibshirani, R. *Regression shrinkage and selection via the lasso*. J. Royal. Statist. Soc B., **58**(1), 267–288 (1996).
- [31] van der Vaart, A. W. *Asymptotic statistics* Cambridge University Press, Cambridge (1998).

## Legends

**Figure 1 : Spike train model using multivariate Hawkes processes.** *Top:* For the 2 presented trains ( $N^{i,j}$ ), each dot indicates the time of occurrence of a spike. The probability that a new spike occurs at time  $t$  on the spike train  $N^i$  is given by the conditional intensity  $\lambda_i(t)$  (see Method). *Bottom:* The function  $h_{j \rightarrow i}$  is defined as piecewise constant on a partition of  $K$  bins of size  $\delta$ . In the illustrated example, 3 preceding spikes occurring on  $N^j$  at different delays (1,2,3) will condition spike generation on  $N^i$  according to the corresponding  $a_{j \rightarrow i}^k$  coefficients (see crosses).

**Figure 2 : Reconstructing a simple Hawkes model network.** **A.** *Left:* Connectivity graph of the simulated network. Each Hawkes process generates an intrinsic firing activity around 11Hz. Hawkes processes 2 and 4 shared a strong common excitatory input from Hawkes process 1 (see Results for details). *Right:* The % of connectivity graphs that were perfectly reconstructed when analyzing 100 datasets of a given duration are plotted for different values of  $\gamma$ . **B.** Simulated (red lines) and typical reconstructed functions  $h_{j \rightarrow i}$  obtained with datasets of 50s (green lines) and 360s (black and blue lines) duration. The optimization was performed using  $\gamma = 3$  and either  $K=10$ ,  $\delta=5\text{ms}$  (green and blue lines) or  $K=20$ ,  $\delta=2\text{ms}$  (black lines). Note that only the functions  $h_{j \rightarrow i}$  of the 2 existing excitatory connections and the 2 potential false additional connections are presented for clarity.

**Figure 3 : Detection of a single connection within a small network.** **A.** Connectivity graph of the simulated networks. Each LIF received a Poisson excitatory input (not shown) generating intrinsic firing rate around 11Hz. Simulations were run while systematically increasing the weight of the excitatory ( $w_e$ ) or inhibitory ( $w_i$ ) connection between LIFs 5 and 8. **B.** Color coded graphs presenting the minimal  $\gamma$  values ( $\gamma_{lim}$ ) that guaranteed an additional false interaction in less than 5% of the datasets as a function of the excitatory (left graph) or inhibitory (right graph) weight and the duration of the datasets. **C.** Graphs presenting the % of datasets, for which the network connectivity with the single excitatory (**1** : top graphs) or inhibitory (**2** : bottom graphs) input from LIFs 5 to 8 was perfectly estimated (TDR), as a function of both dataset duration and connectivity weight.

**Figure 4 : Typical  $h_{j \rightarrow i}$  functions of single excitatory and inhibitory connection between LIFs.** **A.** Reconstructed  $h_{j \rightarrow i}$  function of the excitatory connection (weight=0.6) within the network presented in Figure 3. The estimation was performed using  $\gamma = 3$  and either  $K=10$ ,  $\delta=5\text{ms}$  (red lines) or  $K=20$ ,  $\delta=2\text{ms}$  (black lines) on a dataset of 360s. **B.** Reconstructed  $h_{j \rightarrow i}$  function of the inhibitory connection (weight=-0.6) within the network presented in Figure 3. The estimation was performed using  $\gamma = 3$  and either  $K=10$ ,  $\delta=5\text{ms}$  (red lines) or  $K=12$ ,  $\delta=10\text{ms}$  (black lines) on a dataset of 360s.

**Figure 5 : Robustness of excitatory connection detection.** **A.** *Left :* each of the 4 unconnected LIFs received Poisson excitatory inputs (not shown) to induce spontaneous firing. Simulations were run with various intrinsic firing rates for the 1st LIF ( $\nu_1$  from 11 to 234Hz, non-regular grid) while the 3 other LIFs were kept at a constant firing rate of 11Hz. *Right :* Color coded graph presenting  $\gamma_{lim}$  the minimal values for  $\gamma$  that guaranteed an additional false interaction in less than 5% of the datasets as a function of the 1st LIF firing rate and the duration of the datasets. In simulations where none of the tested values for  $\gamma$  could fulfill the required condition the corresponding squares were left white. **B.** *Top Connectivity graphs:* Each LIF received a Poisson excitatory input (not shown) generating intrinsic firing rate around 11Hz. Simulations were run

while systematically increasing the weight of the excitatory ( $w_e$ ) connections between the LIFs  $1 \rightarrow 2$  and  $2 \rightarrow 4$ . *Top color coded graph:* Color coded graph presenting  $\gamma_{lim}$  the minimal values for  $\gamma$  that guaranteed an additional interaction in less than 5% of the datasets as a function of excitatory weight and duration of the datasets. In simulations where none of the tested values for  $\gamma$  could fulfill the required condition, the corresponding squares were left white. *Middle color coded graph:* same graph constructed from models where connections with *energy* inferior to 0.13 were discarded. *Bottom color coded graph:* same graph constructed from models where connections were discarded according to the "first large jump correction" (see *Results*). **C.** Same legends as in **B** for 4 LIF networks where LIFs 2 and 4 shared a common input (top Connectivity graph).

**Figure 6 : Small network implementing common excitatory inputs. A.** Connectivity graph of the simulated networks. Each LIF received a Poisson excitatory input (not shown) generating intrinsic firing rate around 11Hz. LIFs 2 and 4 shared a common input of weight  $w_e$  from LIF 1. **B.** Color coded graphs presenting the % of 60s long datasets where the presence of a connection from  $LIF_j$  to  $LIF_i$  (non null  $h_{j \rightarrow i}$ ) was indicated by Hawkes models fitted with  $\gamma = 3$ . For  $w_e=0.25$  (top graph) only the auto-interaction functions,  $h_{i \rightarrow i}$ , and the 2 functions  $h_{1 \rightarrow 2}$  and  $h_{1 \rightarrow 4}$ , corresponding to existing connections within the network were reported. However, when increasing  $w_e$  to 0.60 (bottom graph), Hawkes models suggested the presence of a non-existing bidirectional connection between LIFs 2 and 4 in less than 10% of the datasets. **C.** Color coded graphs of the "energy" of the  $h_{j \rightarrow i}$  functions for simulations with  $w_e=0.25$  (top graph) and 0.60 (bottom graph). Note that the average values of the "energy" were calculated from datasets reporting non-null  $h_{j \rightarrow i}$  functions.

**Figure 7 : Discarding false interaction functions using the first large "jump" correction. A.** Connection diagram of the simulated network. The strong excitatory connections of weight  $w_e = 0.60$  are indicated with black edges. Each LIF received a Poisson excitatory input (not shown) generating intrinsic firing rate around 11Hz. The presence of multiple common inputs in this strongly connected network increases the probability that some non-existing connections (red dotted edges) may be added in the reconstructed networks. **B.** Color coded graph presenting the % of 60s long datasets where the presence of a connection from  $LIF_j$  to  $LIF_i$  (non null  $h_{j \rightarrow i}$ ) was indicated by Hawkes models fitted with  $\gamma = 3$ . As expected, non-existing connections were suggested in a small number of cases. **C.** Plots showing the values  $E_{j \rightarrow i} = \int h_{j \rightarrow i}$  sorted in increasing order for 3 datasets of 40s (graph 1.), 60s (graph 2.) and 300s (graph 3.). In each plot, black edges indicate the largest "jump" in the ordered sequence of  $E_{j \rightarrow i}$  and red edges indicate 15% of this largest "jump". Discarding the functions  $h_{j \rightarrow i}$  corresponding to the  $E_{j \rightarrow i}$ 's smaller than this threshold removed the non-existing connections (highlighted in grey) but did not discard any existing connections.

**Figure 8 : Inferring functional connectivity graphs with Hawkes model-based method. A.** Number of additional and missing excitation/inhibition connections (median and 0.1 quantile values) in the reconstructed functional connectivity graphs of 89 randomly generated LIF networks (see *Results*) as a function of dataset duration. Hawkes models were fitted using a  $\gamma$  value of 3. **B.** Left graph : Percentage as a function of the dataset duration of 89 randomly generated LIF networks in which the excitatory (red line) or inhibitory (blue line) connections were perfectly inferred with the Hawkes model-based method. The green line shows the % of non-existing excitation added in the Hawkes models. The dashed line materializes the 5% limit. Right graph : same graph constructed from models where connections were discarded according to the "first large jump method" (see *Results*).

**Figure 9 : Computational cost of the Hawkes model-based method.** Graph presenting the median (black line) and 0.1 quantile values (dotted lines) of the required computational time to reconstruct functional connectivity graphs of 89 randomly generated LIF networks (see Figure 8) as a function of dataset duration (with  $K = 10$ ,  $\delta = 0.005$ ). Parallel computation using 4 threads were performed on a MacBook Pro 2.7GHz with 8GB of RAM.

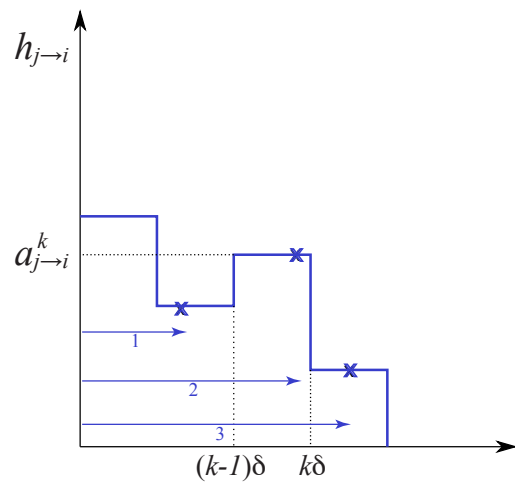
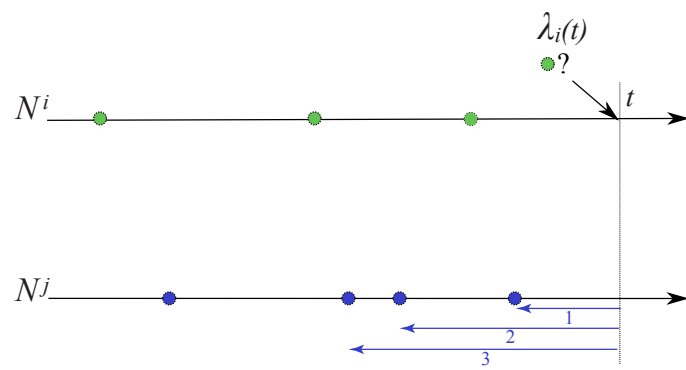


Figure 1:

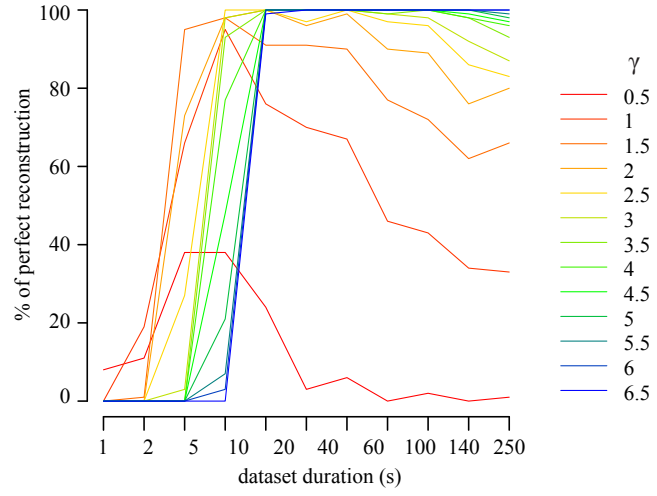
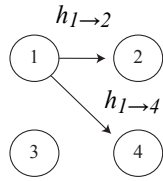
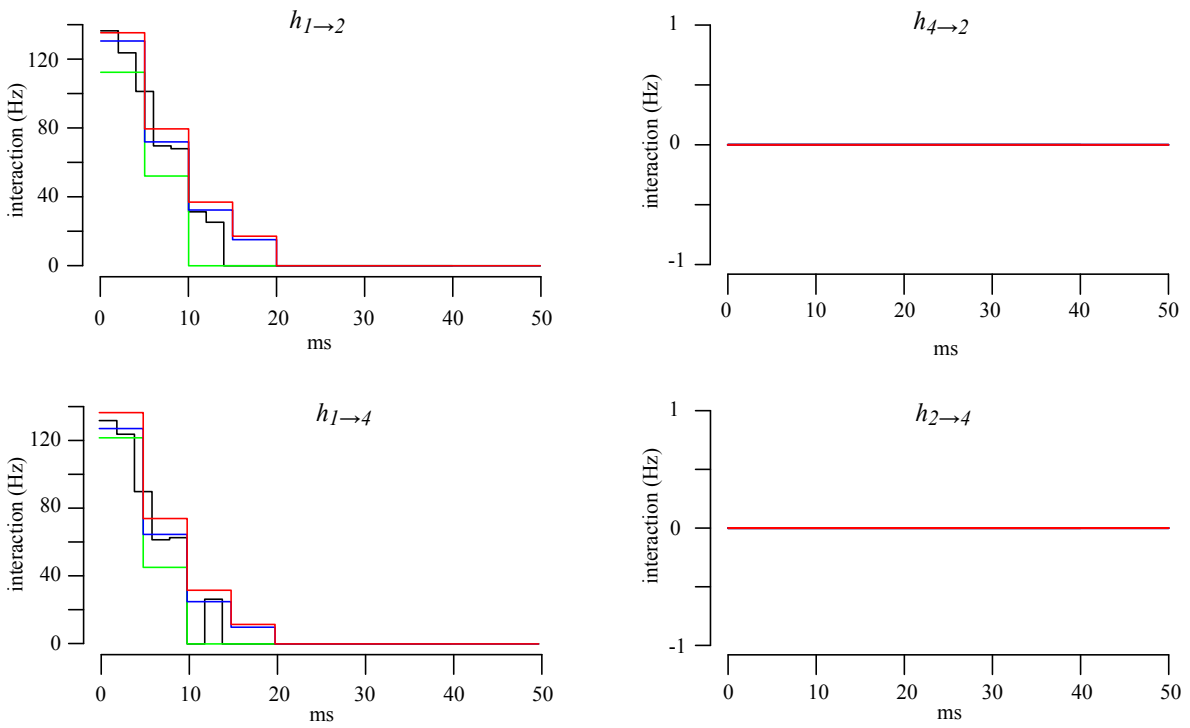
**A****B**

Figure 2:

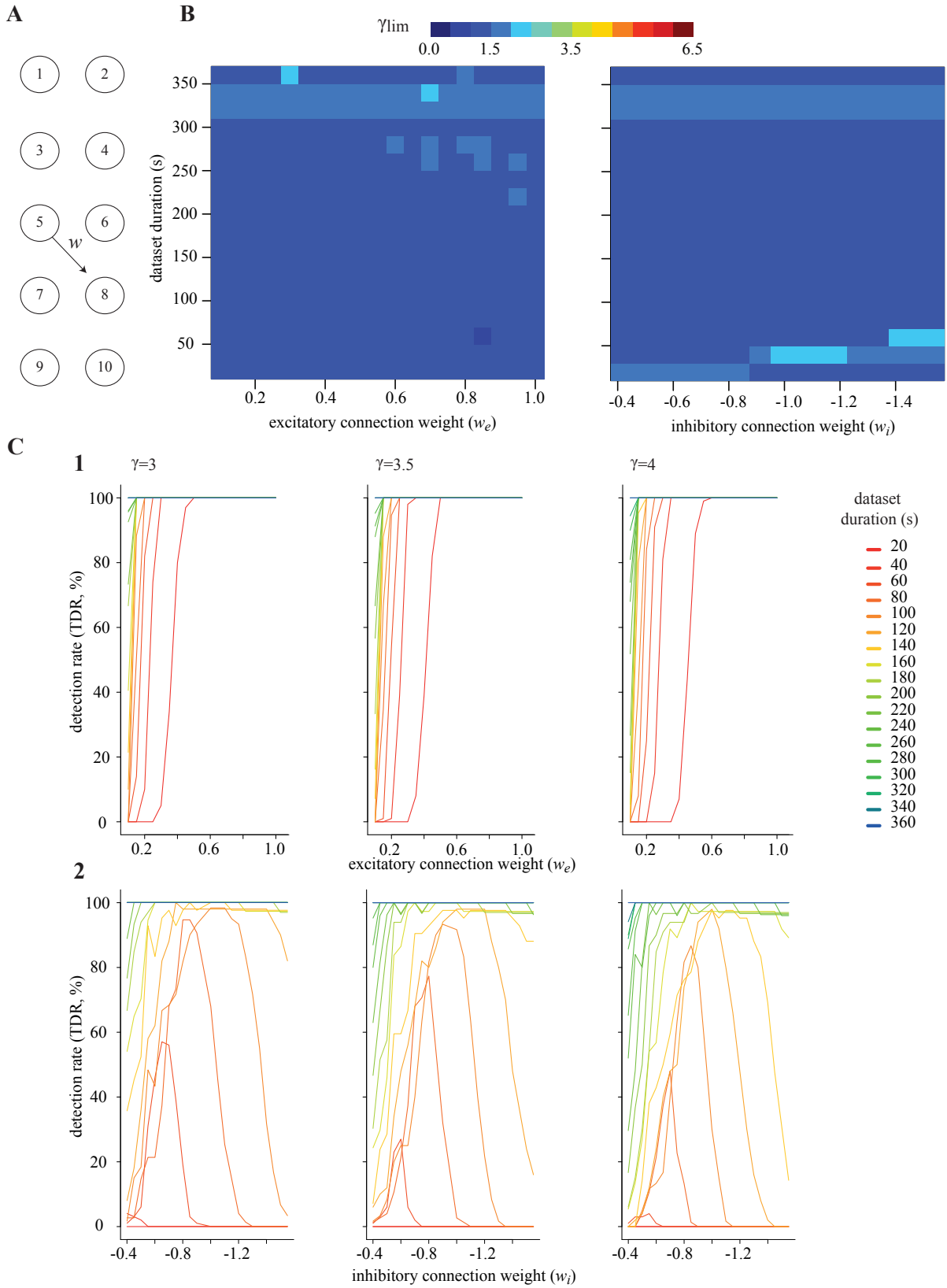


Figure 3:

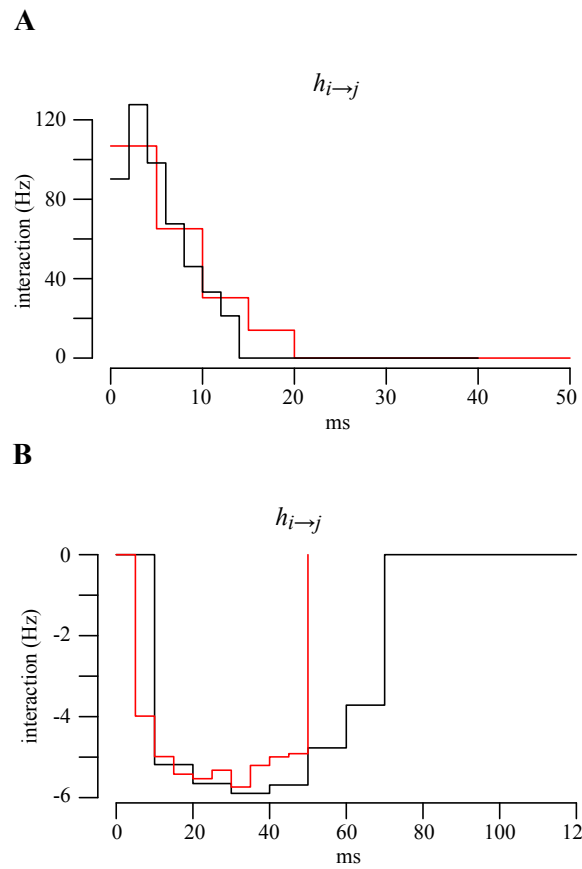


Figure 4:



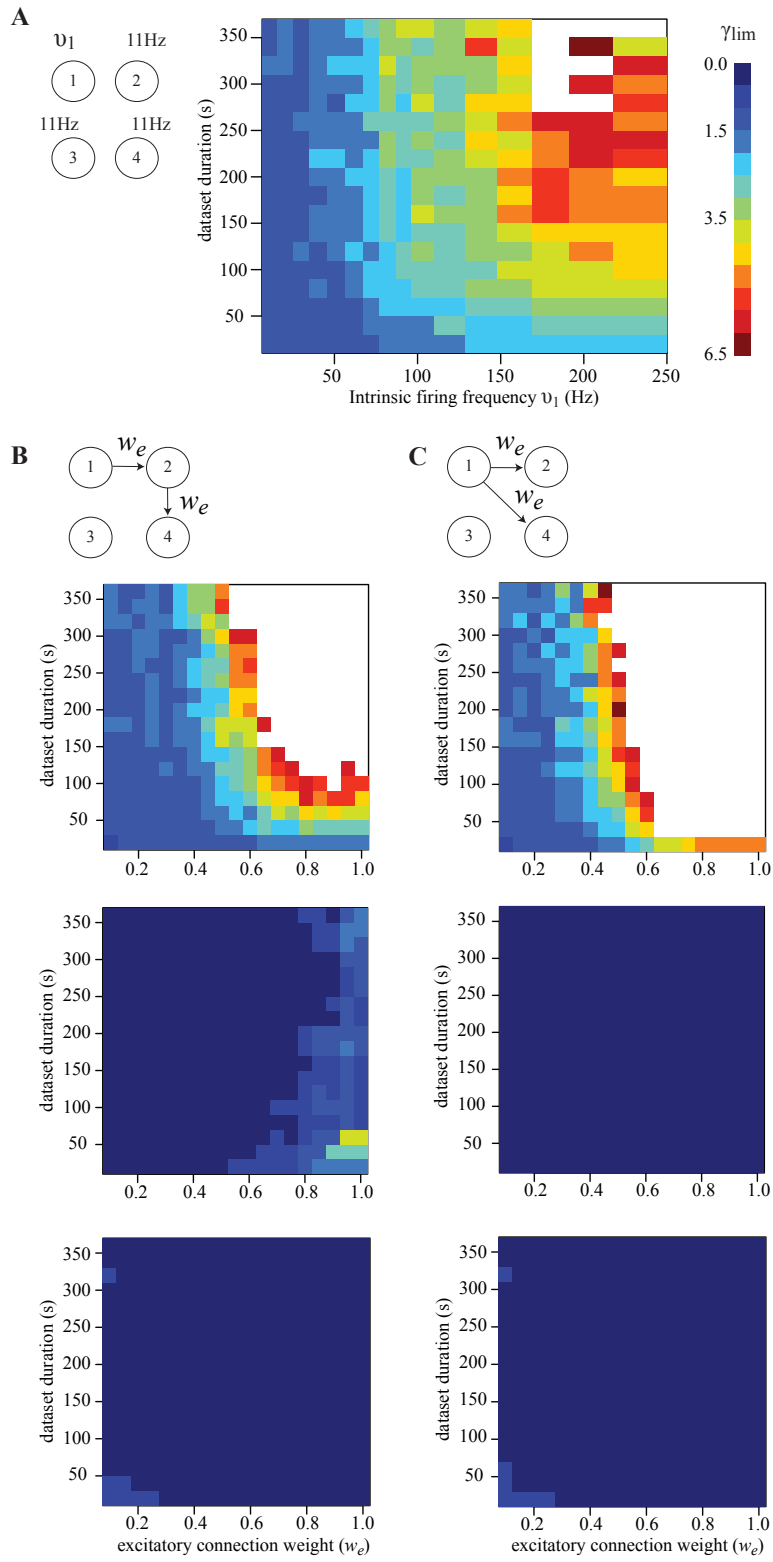


Figure 5:

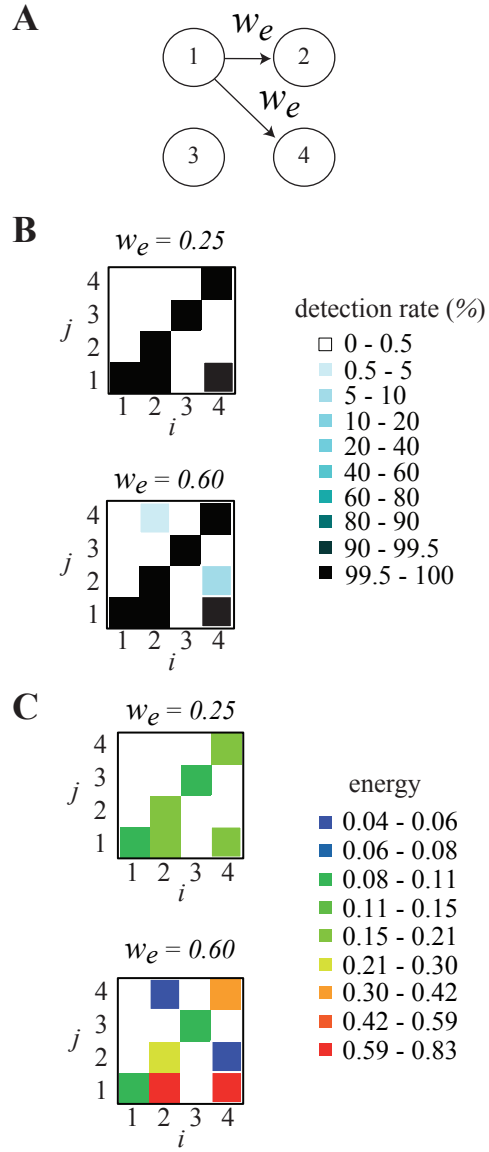


Figure 6:

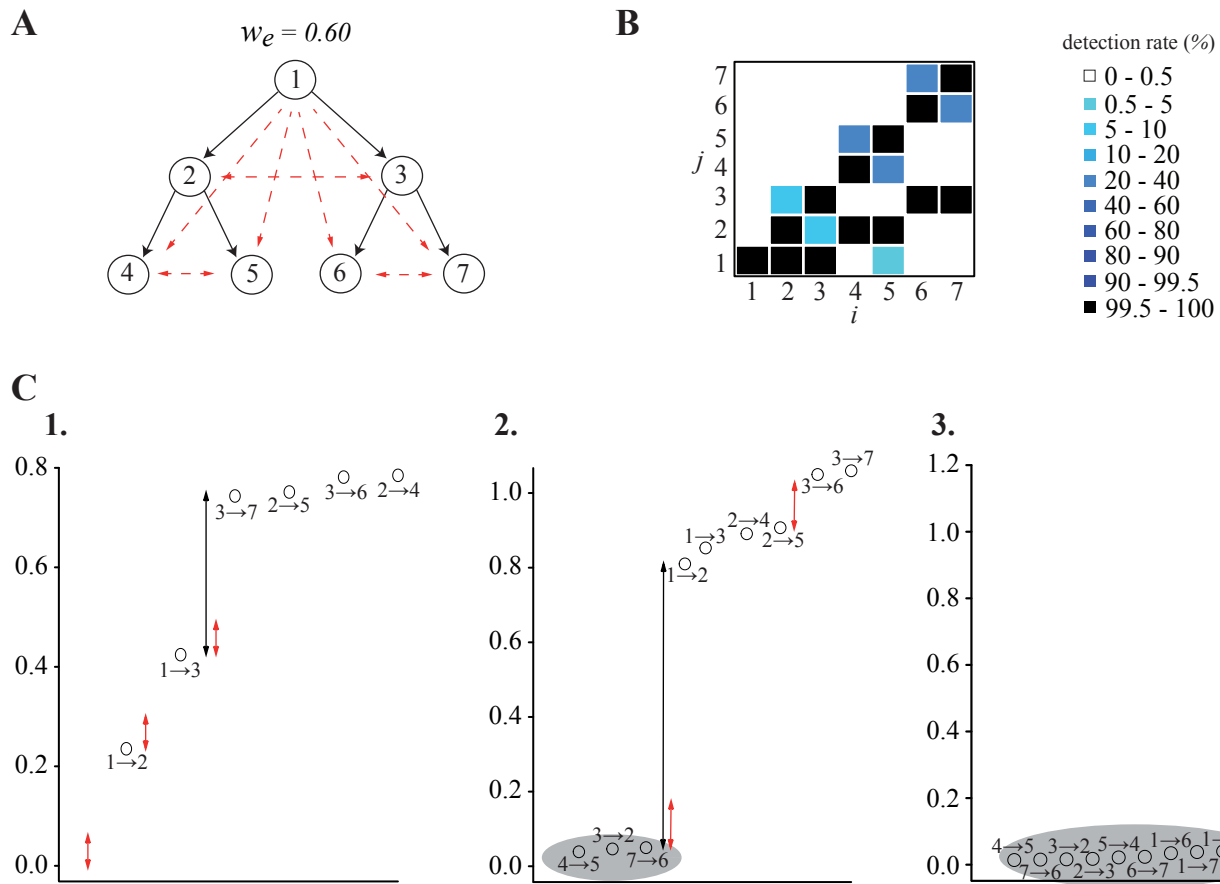


Figure 7:

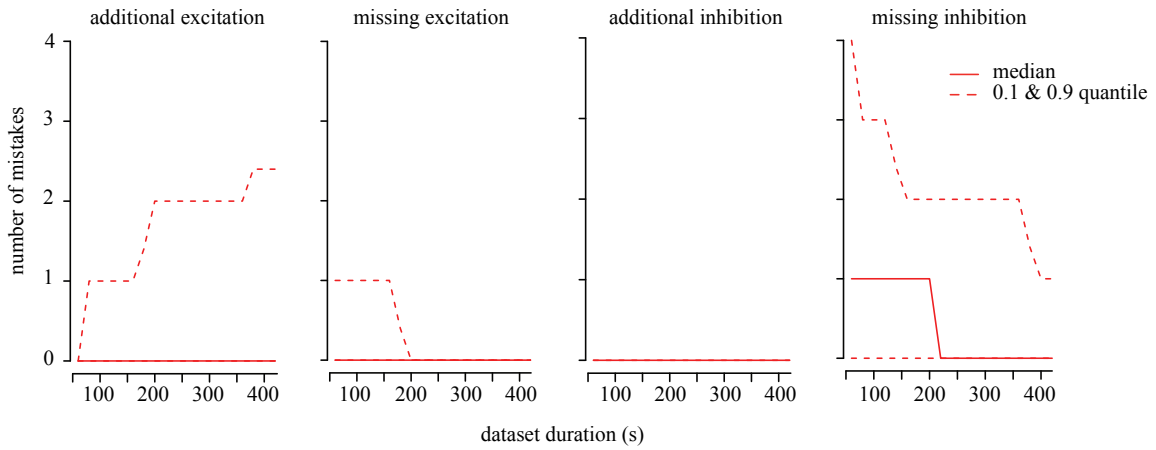
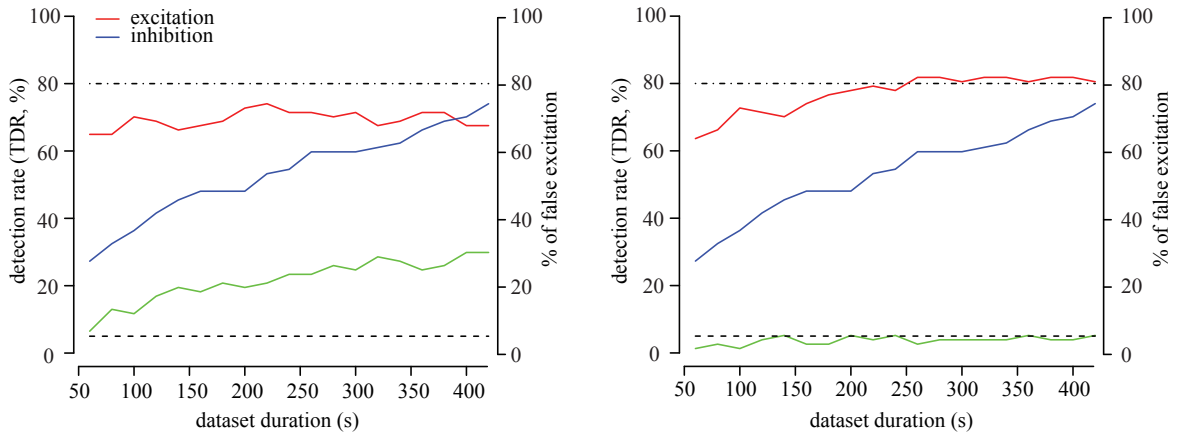
**A****B**

Figure 8:

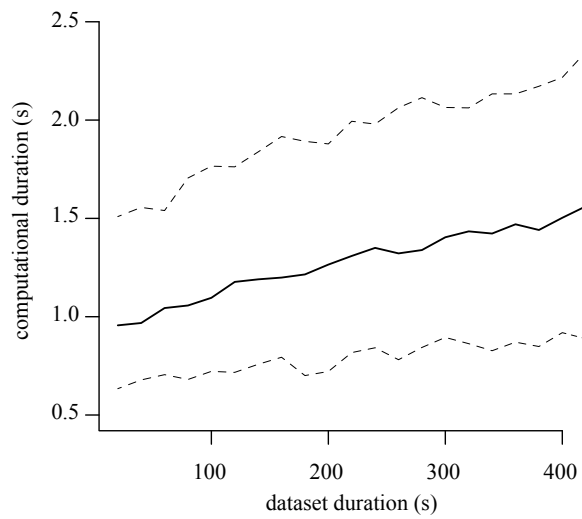


Figure 9: