Discovering Neuronal Connectivity from Serial Patterns in Spike Train Data

Casey O. Diekman[1], **Kohinoor Dasgupta**[2], **Vijay Nair**[2], **K.P. Unnikrishnan**[3]

[1]Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102

[2]Department of Statistics, University of Michigan, Ann Arbor, MI 48109

[3]Center for Biomedical Research Informatics, NorthShore University HealthSystem, Evanston, IL 60201

Keywords: functional connectivity, spike train analysis, temporal data mining, multi-electrode array

Abstract

Repeating patterns of precisely-timed activity across a group of neurons (called frequent episodes) are indicative of networks in the underlying neural tissue. This paper develops statistical methods to determine functional connectivity among neurons based on "non-overlapping" occurrences of episodes. We study the distribution of episode counts and develop a two-phase strategy for identifying functional connections. For the first phase, we develop statistical procedures that are used to screen all two-node episodes and identify possible functional connections (edges). For the second phase, we develop additional statistical procedures to prune the two-node episodes and remove "false" edges that can be attributed to chains or fan-out structures. The restriction to non-overlapping occurrences makes the counting of all two-node episodes in phase one computationally efficient. The second (pruning) phase is critical since phase one can yield a large number of false connections. The scalability of the two-phase approach is examined through simulation. The method is then used to reconstruct the graph structure of observed neuronal networks, first from simulated data and then from recordings of cultured cortical neurons.

1 Introduction

Multi-electrode array recordings have become a standard tool in neuroscience, leading to large collections of neuronal spike train data. However, there has not been an analogous development in statistical methods to extract the spatial and temporal information in these large data sets quickly and efficiently [7]. The goal of this paper is to develop statistical methods for analyzing functional connectivity among multi-neuronal spike train data and use them to reconstruct the graph structure of the underlying neuronal network.

The primary contribution of the paper is the development of a two-phase strategy and associated statistical procedures for identifying functional connections based on "frequent episodes" in spike train data. In phase one, the methodology counts the occurrences of all two-node episodes and screens them to identify possible functional connections (edges). In phase two, the set of identified two-node episodes is pruned to remove the potentially large number of "false" edges that can be attributed to chains or fan-out structures. In this paper, we focus on episode counts based on the concept of "non-overlapping occurrences." The restriction to non-overlapping occurrences is important for computational reasons. Computationally-efficient algorithms have been developed in the literature [20], [21], and [27]. However, the statistical theory associated with non-overlapping counts has not been fully developed [9, 33]. The paper develops this theory and uses it to derive associated statistical procedures for phases one and two. The two-phase approach is then applied to spike train data in order to reconstruct the graph structure of both simulated and experimentally observed networks of neurons.

Counts of serial episodes are statistical summaries that characterize the directional relationships among neurons. They are formally defined in Section 2, where we also provide additional background for the data types considered in the paper. Section 3 develops the distribution theory for serial episodes under a stochastic framework. The results are then used in Section 4 to develop statistical thresholds for determining the important episodes (those that correspond to strong functional connectivity among the neurons) and to estimate the strengths of the connectivity. In Section 5, we show how the results can be used to recover the structure of simulated networks and examine the scalability of the method. Section 6 illustrates application of the methods to multi-electrode array data from cortical cultures. The paper concludes with Section 7 discussing how the results can be extended.

2 Background

Spike train data are time-ordered sequences of events, as indicated by the illustrative data set involving five neurons A–E shown in Table 1. In this case, neuron A fired at time unit 1, neuron D fired at time unit 3, and so on. When a neuron influences another, the influence is mediated by processes (e.g. synaptic communication) that have characteristic time delays on the order of milliseconds. Thus, the difference between the times of the two spikes should differ by some time unit k. The subset $< (A, 1), (C, 4) >$ in Table 1 indicates C occurring $k = 3$ milliseconds after A. Temporal patterns in spike train data (such as the firing of one neuron followed by the firing of another after a fixed time) that occur frequently are indicative of neuronal microcircuits. Automatically detecting such patterns, and using statistical methods to separate patterns due to underlying connections from those that occur randomly, is very useful for understanding the functional connectivity in the neural tissue.

[(A,1), (D,3), (C,4), (A,5), (B,7), (C,8), (B,11), (C,12), (E,12), (A,13), (B,15), (C,16)]

Table 1: An Illustrative Example of Spike Train Data for Five Neurons

Serial episodes are temporal patterns with a specific ordering of the events, while *parallel episodes* have no constraints on the relative order of events in the pattern [22]. The analysis in this paper focuses on serial episodes with inter-event time constraints. For an episode involving the two neurons A and B, we represent the time information through the notation $A[k]B$, where k is the inter-event delay. For example, in Table 1, there are three distinct occurrences of the episode $A[3]C$: $< (A, 1), (C, 4) >$; $< (A, 5), (C, 8) >$, and $< (A, 13), (C, 16) >$. A serial episode that occurs a "large" number of times in the data is called a frequent episode. Clearly, the term "large" needs further quantification, and this is in fact one of the goals of the paper. Further, we will restrict attention to two-node interactions in this paper. We interpret frequent multi-node episodes of the form $A[k_1]B[k_2]C$ as indicative of either chain structure (consecutive occurrences of two-node episodes such as $A[k_1]B$ followed by $B[k_2]C$), fan-out structure (co-occurrence of the two-node episodes $A[k_1]B$ and $A[k_1 + k_2]C$), or both (feed-forward loop structure). See Figure 2 and [3, 23]. Chains and fan-out structures are known to lead to systematic prediction errors (in particular false positives) in algorithms for inferring gene regulatory networks [23]. The procedures developed in this paper are designed to infer these structures accurately. There is some discussion in the neuroscience literature about "true" third- and higher-order interactions [24, 4, 12]. However, [34] suggests that a significant amount of the connectivity in neuronal networks can be explained by just two-node interactions.

The time delay for a connection between two neurons will in general not be known, so we will count the occurrences of episodes for a range of possible values of k. This in turn implies that we need computationally efficient algorithms for counting. Frequent episode discovery, a general method in temporal data mining [22], has been proposed as a method for efficiently counting the numbers of serial episodes with inter-event time constraints [27]. It is also possible that the time delay is random. While the methods in this paper can be adapted to handle this situation, we will not pursue it here.

[(A,1), (B,2), (A,3), (A,5), (B,6), (B,8), (A,9), (B,10), (A,12), (B,14)]

Table 2: Illustration of Non-overlapping Occurrences

The analysis in this paper will focus on episode counts based on "non-overlapping" occurrences. To explain this concept, consider the sequence for two neurons A and B in Table 2. There are several occurrences of the episode $A[5]B$ in this sequence: <(A,1),(B,6)>, <(A,3),(B,8)> and <(A,5),(B,10)>. The last two occurrences *overlap* with the first in the sense that the events (A,3) and (A,5) occur before (B,6). In this case, we cannot determine whether the firing of (B,6) was influenced by (A,1), (A,3) or (A,5). If we are to interpret a serial episode as (part of) a microcircuit in the underlying neural tissue, then it should represent a series of causative or triggering events. A frequent episode based on counts of non-overlapping occurrences would be such a series. To allow for this, some researchers [2] restrict attention to only non-overlapping occurrences, namely occurrences with non-overlapping time periods. Figure 1 summarizes the number of overlapping and non-overlapping occurrences for the episode $A[5]B$ for a longer data sequence. There are $N = 6$ counts of $A[5]B$ and only $M = 3$ of these are non-overlapping occurrences.

An additional advantage of restricting attention to non-overlapping occurrences is that it is more efficient to count the number of episodes – an important concern when dealing with large data sets and searching for episodes at all possible values of k, the time delays. Efficient algorithms have been developed for such situations by [20], [21], and [27]. These algorithms use finite-state automata, and part of their efficiency comes from the fact that to count non-overlapping occurrences only one automaton per episode is needed.

The rest of the paper is organized as follows. Section 3.1 describes the basic model and develops the distributions of the number of occurrences of episodes. The main contribution here is the development of the distribution based on non-overlapping occurrences. In [33], recurrence relations are used to compute just the mean and variance of this random variable. Our results are more general and we provide a complete characterization that allows us to compute the

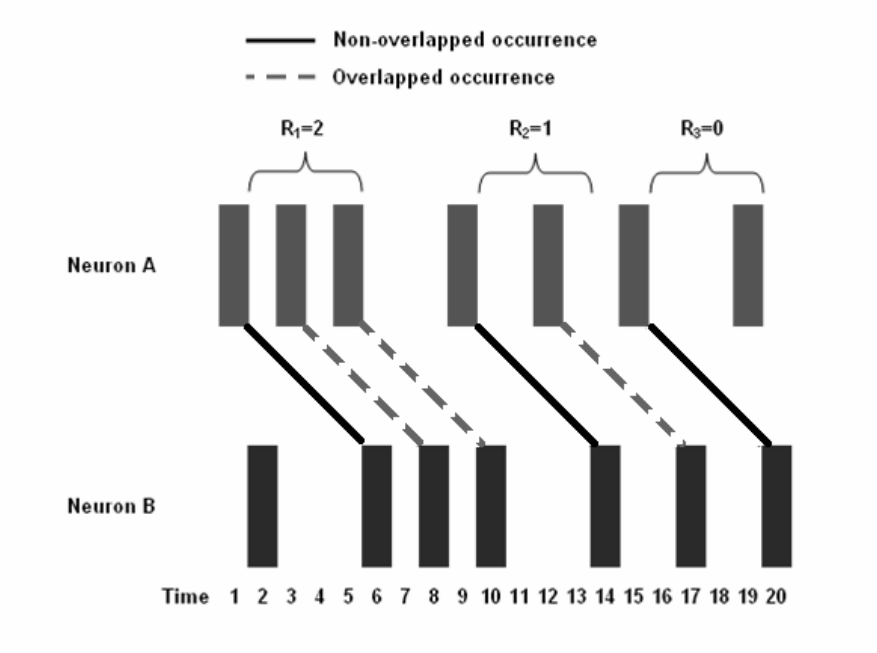

Figure 1: Non-overlapping occurrences of episode A[5]B. The total number of occurrences in $L = 20$ bins is $N = 6$, consisting of $M = 3$ non-overlapping occurrences and $R_1+R_2+R_3 = 3$ overlapping occurrences.

moment generating function and, hence, all moments. The results also allow us to compute the covariances of the random variables for different types of episodes, which are needed for testing various hypotheses of interest. We use the distributional results to develop statistical thresholds for separating the observed episodes into two groups: a) random (arising from pairs of nodes that are independent), and b) active (pairs of nodes that are connected). A simple approach is to test for independence of the pairs of nodes. However, in practice, researchers are interested in detecting only sufficiently strong connections (as indicated by some measure of strength). Section 4 considers an appropriate notion of strength and develops methods for determining connectivity above (or below) the specified measure of strength. Two issues arise in implementing the methods: a) screening out false positives that occur due to chains and fan-outs; and b) controlling the overall error rates due to multiple testing. Both of these topics are also discussed in Section 4. The results are then applied to simulated networks in Section 5 and data from cortical cultures in Section 6. Section 7 briefly discusses extensions to non-stationary models and other practical issues.

3 Distributions of Episode Counts

3.1 The Model

Multivariate point processes (and their discrete versions) are often used to model multi-neuronal spike trains. See [6, 8, 38, 31] and references therein. In this paper, we consider a class of dependent Bernoulli processes to characterize the behavior of a set of neurons. These can be viewed as discretized versions of the Poisson processes that have been extensively discussed in the literature.

Let J be the number of neurons being observed. Let Δ be a small enough interval so that there is at most one spike per neuron in the interval. We observe the process over a period of L consecutive intervals. Let $X_j(t) = 1$ if neuron j fires at time t and zero otherwise, for $t = 1, \ldots, L$ and $j = 1, \ldots, J$. Let P_j be the probability of neuron j firing in an interval and denote the number of firings during L intervals as N_j for $j = 1, \ldots, J$. In the artificial case where the neurons fire independently, the $X_j(t)$'s follow independent Bernoulli processes, and N_j's will be independent binomial(L, P_j) random variables.

We are interested in the case where there is dependence among the neurons, and we formulate it as follows. Suppose we have two neurons A and B with unidirectional dependence: A influences B, but B does not influence A, and there is a delay of k time units between the time A fires and when it affects the probability of B firing (written as $A \xrightarrow{k} B$). Since the firing of A is unaffected by B, $X_A(t)$ is still a Bernoulli process with success probability P_A. The firing of neuron B is due to a mixture of two processes – it can fire on its own, or it can fire due to excitation by A. Since each of these processes are independent over time, the firing of B across different intervals remains independent. The mixture probability of B firing in any interval t is:

$$\begin{aligned} P(X_B(t) = 1) &= P(X_B(t) = 1 | X_A(t-k) = 1) P(X_A(t-k) = 1) \\ &+ \; P(X_B(t) = 1 | X_A(t-k) = 0) P(X_A(t-k) = 0) \\ &= P_A P_{B|A} + (1 - P_A) P_{B|\bar{A}} \\ P(X_B(t) = 0) &= 1 - P(X_B(t) = 1) \end{aligned}$$

Further, $X_B(t)$ is marginally a Bernoulli process with $P_B^c = P_A \, P_{B|A} + (1 - P_A) \, P_{B|\bar{A}}$, where the superscript c stands for combined. However, $(X_A(t), X_B(t))$ are now dependent.

While the above formulation is relatively simple, it has been found to be reasonable for modeling the firing behavior of cortical neurons [35]. Nevertheless, neurons do exhibit time-varying behavior, so the reasonableness of the Bernoulli model for a particular dataset of interest should be tested empirically [16]. If the temporal behavior is piece-wise stationary or slowly

varying, the results can be applied to segments of time intervals within which the behavior is approximately stationary. Another issue is that, immediately following a spike, a neuron may exhibit a short (typically less than 1 ms) absolute refractory period during which it cannot spike again [15, 25]. Such refractory effects become detectable when neurons have high enough firing rates.

3.2 Distribution in the Unrestricted Case

Consider first the case where all occurrences of an episode are counted, i.e., attention is not restricted to non-overlapping occurrences. Suppose the episode of interest is $E = A[k]B$: a firing of neuron A followed k seconds later by a firing of neuron B. For $t = 1, ..., L-k$, consider the binary process $I_E(t)$ that equals 1 if $X_A(t) = 1$ AND $X_B(t+k) = 1$. The $I_E(t)$'s are iid Bernoulli with success probability $P_{A[k]B} = P_A P_{B|A}$ for $t = 1, ..., L-k$. Let N denote the number of times that $I_E(t)$ equals 1, i.e., the number of occurrences of the serial episode E. The possible values of N range from 0 to $(L-k)$; if A fires after time $t = L-k$, an occurrence of the episode would not be complete by time $t = L$. It is clear that N has a binomial distribution with parameters $(L-k)$ and $P_{A[k]B}$. Therefore, one can estimate $P_{A[k]B}$ using the simple estimator $\hat{P}_{A[k]B} = N/(L-k)$. $\mathrm{Var}(\hat{P}_{A[k]B}) = P_{A[k]B}(1-P_{A[k]B})/(L-k)$. This result holds provided there are no cycles among the neurons, i.e., the network structure connecting the neurons is a directed acyclic graph (DAG). If neuron A influences B and B in turn influences A, the model becomes much more complicated. We do not consider such situations in this paper.

3.3 Distribution in the Non-Overlapping Case

The technically challenging case is the distribution of serial episode counts based on non-overlapping occurrences. Consider again two neurons A and B, and let M be the number of *non-overlapping* occurrences of the episode $E = A[k]B$. Iterative methods for calculating the mean and variance of M were developed in [33]. We provide a complete characterization of the distribution and also obtain an expression for its moment generating function (and hence moments of all orders).

We can partition N, the total number of occurrences, as follows. Consider the j^{th} non-overlapping occurrence of E, and let R_j be the number of (unrestricted) occurrences that occur between the $(j-1)^{th}$ and j^{th} non-overlapping occurrences (see Fig. 1 for an illustration). Then, it is easily seen that $N = M + \sum_{j=1}^{M} R_j$. Further, the R_j's and M are independent of each other. So, we have the following result:

Proposition: N has the same distribution as

$$N \;=\; M + \sum_{j=1}^{M} R_j \tag{1}$$

Further, the R_j's are iid binomial($k, P_{A[k]B}$) random variables.

This provides a complete characterization of the distribution of M. While we cannot use the result to compute the distribution of M analytically, the moment-generating function (and hence all the moments of M) can be computed easily. The mean and variance can be obtained directly. Taking expectations of both sides of Equation 1, and using the independence of the R_j's and the distribution of N, we have

$$\mathrm{E}\left[M\right] \;=\; \frac{\mathrm{E}\left[N\right]}{1+\mathrm{E}\left[R_1\right]} = \frac{L-k}{1/P_{A[k]B}+k} \tag{2}$$

Similarly, the variance of M is obtained as

$$\begin{aligned} \mathrm{Var}[M] &= \frac{(L-k)P_{A[k]B}(1-P_{A[k]B})}{(1+kP_{A[k]B})^3} && (3)\\ &= \frac{\mathrm{Var}[N]}{(1+kP_{A[k]B})^3} && (4) \end{aligned}$$

To get the moment generating function (MGF) of M, we write

$$\begin{aligned} \mathrm{E}\left[\exp(tN)\right] &= \mathrm{E}\left[\exp(t(M+\sum_{j=1}^{M} R_j))\right] \\ &= \mathrm{E}_M[\mathrm{E}_X[\exp(t(M+\sum_{j=1}^{M} R_j))|M]] \\ &= \mathrm{E}\left[\exp(tM)(Q_{A[k]B} \;+\; P_{A[k]B}\;\exp(t))^{Mk}\right] \\ &= \mathrm{E}\left[\exp(M(t+klog(Q_{A[k]B}+P_{A[k]B}\exp(t))))\right] \qquad (5) \end{aligned}$$

where $Q_{A[k]B} = 1 - P_{A[k]B}$. Let $s(t) = t + k\log(Q_{A[k]B} + P_{A[k]B}\exp(t))$. This is a one-to-one and strictly increasing function of t, so a unique inverse exists. Denote this inverse function by $g(s)$. (Note that $g(s)$ has to be obtained numerically.) Then, the MGF of M can be expressed as

$$\mathrm{E}\left[\exp(sM)\right] = \mathrm{E}\left[\exp(g(s)N)\right] = (Q_{A[k]B} + P_{A[k]B}\exp(g(s)))^{L-k}.$$

We can get the n^{th} moment of the distribution of M by differentiating the MGF n times w.r.t. to s and evaluating it at $s = 0$.

The distribution of M (CDF or pmf) can be approximated up to any degree of precision by Monte Carlo sampling and the characterization in Equation 1. We also examined more direct approximations and found a fourth-order polynomial function of a normal random variable to

be reasonable. Specifically, one can approximate the distribution of M by $\sum_j a_j Z^j$, where Z is a standard normal random variable and the a_j's are chosen to match the first four moments of M [11, 41].

Equation (2) yields an estimator of $P_{A[k]B}$ based on M as

$$\hat{P}^M_{A[k]B} = \Big(\frac{L-k}{M} - k\Big)^{-1}, \tag{6}$$

a non-linear function of M. Its variance can be approximated using Taylor series as

$$\mathrm{Var}[\hat{P}^M_{A[k]B}] \approx \frac{(1+kP_{A[k]B})P_{A[k]B}(1-P_{A[k]B})}{(L-k)}. \tag{7}$$

As a side issue, one might wonder about the loss in efficiency in basing the analysis on non-overlapping occurrences M rather than N. A comparison of the (asymptotic) variances of $\widehat{P}^N_{A[k]B}$ (estimator based on N) and $\widehat{P}^M_{A[k]B}$ (based on M) provides insight into this question. The ratio of the asymptotic variances of $\widehat{P}^N_{A[k]B}$ and $\widehat{P}^M_{A[k]B}$ is

$$\text{Relative Efficiency} = \frac{1}{1+kP_{A[k]B}}. \tag{8}$$

It is independent of L and decreases as k or $P_{A[k]B}$ get large, suggesting a greater loss in efficiency in using M in such situations. Of course, these are also precisely the cases where the number of occurrences will be large, and therefore, are the situations where the gains in computational efficiency due to using M over N will be greatest.

4 Detecting Active Connections

Consider again two neurons A and B. If an excitatory connection is present, $P_{A[k]B} > P_A \times P_B$; for an inhibitory connection, $P_{A[k]B} < P_A \times P_B$. We will restrict attention to the excitatory case as the arguments for the inhibitory case are analogous. The simplest question to ask is if the number of observed occurrences of an episode (based on N or M) is larger than what one would expect under "randomness". This can be formulated as testing the null hypothesis $P_{A[k]B} = P_A \times P_B$ against the alternative $P_{A[k]B} > P_A \times P_B$. If we let $S = P_{A[k]B}/P_AP_B$, the question becomes testing $H_0 : S = 1$ against $H_1 : S > 1$. The value of S can be viewed as a measure of the connection strength between A and B.

Researchers are typically interested in detecting only connections that are sufficiently strong or "active"; weak connections are often not of interest. Let S_0 be a user-specified threshold for strength. Let $\tau(S) = P_{A[k]B} - SP_AP_B$. Then, detecting $S > S_0$ is equivalent to testing the hypothesis

$$H_0 : \tau(S_0) \le 0 \quad \text{against} \quad H_A : \tau(S_0) > 0 \tag{9}$$

In other words, we decide in favor of $S > S_0$ only if the null hypothesis is rejected, a decision that requires sufficiently strong evidence.

4.1 Test Procedures

Consider first the case where one observes N, the episode count based on all occurrences of $A[k]B$. Then, we can estimate the parameters in $\tau(S)$ based on the underlying binomial data. Denote the corresponding estimator as $\hat{\tau}_N(S)$. The variance can be expressed as

$$\mathrm{Var}(\hat{\tau}_N(S)) = \mathrm{Var}(\hat{P}_{A[k]B}) + S^2\,\mathrm{Var}(\hat{P}_A\hat{P}_B) - 2S\,\mathrm{Cov}(\hat{P}_{A[k]B}, \hat{P}_A\hat{P}_B). \tag{10}$$

Expressions for these components are obtained in the Appendices A.1 and A.2. The problem is non-trivial since $\hat{P}_A$ and $\hat{P}_B$ are not independent when $S > 1$. The estimated variance of $\hat{\tau}_N(S)$ can be obtained by substituting the estimates for the unknown parameters in the variance expressions in Appendix A.1. One can then test the hypothesis of an active connection given in Equation (9) using a normal approximation for the standardized statistic $Z_\tau = \hat{\tau}_N(S_0)/\sqrt{\widehat{\mathrm{Var}}(\hat{\tau}_N(S_0))}$. In practice, one can implement the test procedure by computing the approximate lower confidence bound corresponding to the desired α-level for $\tau(S_0)$ and checking if the confidence bound does or does not include the value zero.

The more technically challenging problem is to develop analogous procedures based on M or equivalently $\hat{\tau}_M(S_0)$. While formulation of the problem and the structure of the estimator $\hat{\tau}_M(S_0)$ are the same as before, the computation of $\mathrm{Var}(\hat{\tau}_M(S_0))$ is now much more involved. Knowledge of the third and fourth order moments of M, discussed in Section 3.3, is needed for this purpose. The variance expressions are developed in Appendices A.3 and A.4. An illustration of these procedures is discussed in Section 5. We have examined the adequacy of the normal approximations to the test statistics through simulation studies and found them to be quite adequate.

4.2 Multiple Testing

The proposed procedure involves testing all pairs of nodes at all possible values of the delay k to identify the active two-node episodes. When either the number of neurons or the number of possible values of k is large, one would have to be concerned about the problem of multiple testing. Performing hundreds or thousands of tests can lead to the spurious detection of false positives. There are several methods in the literature for handling multiple comparisons. These include Bonferroni correction to control the overall error, and the more recent technique of false discovery rates or FDRs (see, for example [36, 5]). We used simulation studies to determine the extent

of false positives from multiple testing and which methods for multiple comparisons perform well in our context. We found the Bonferroni correction to be too conservative, as is well-known to be the case in the literature. The FDR method is not ideally suited for our situation and did not perform particularly well either. The approach we found to be most effective in our simulation studies was to use S_0, the user-specified threshold for strength, as a tuning parameter. As we increased the value of S_0 from one (null hypothesis of independence) to about 3 or 4, most of the false positives due to multiple testing disappeared (see Table 3). The statistical methods, as expected, perform well in detecting connections with relatively strong strengths. Therefore, we propose the use of the strength threshold as a tuning parameter: try different values of S_0 and select the value where the conclusions (number of edges detected) stabilize.

4.3 Eliminating False Edges

In addition to false positives due to multiple testing or Type I error, certain types of network structure can lead to systematic false positives [23]. We refer to false positives due to network structure as *false edges*. Figure 2 shows the two kinds of false edges that are likely to occur (the dashed lines). The left panel shows the situation where we have a pair of active connections $A[k_1]B$ and $B[k_2]C$. This may lead to the detection of $A[k_1 + k_2]C$ as statistically significant even if it is not an active connection. The center panel shows a different scenario where false edges can occur. When there are active connections $A[k_1]B$ and $A[k_1+k_2]C$, then $B[k_2]C$ may be detected as active even if it is not. So, we need to do a second pass through the connections that are detected as active to determine if there are false edges due to this phenomenon.

We propose procedures for testing and screening out false edges in the different scenarios described above. Suppose we have detected all three of the episodes $A[k_1]B$, $A[k_1 + k_2]C$, and $B[k_2]C$ as significant. Consider first the case in the left panel of Figure 2 and the events $AC = A[k_1+k_2]C = \{X_A(t) = 1, X_C(t+k_1+k_2) = 1\}$, $ABC = \{X_A(t) = 1, X_B(t+k_1) = 1, X_C(t + k_1 + k_2) = 1\}$, and $A\bar{B}C = \{X_A(t) = 1, X_B(t + k_1) = 0, X_C(t + k_1 + k_2) = 1\}$ for $t \in [0, L - k_1 - k_2]$. Denote the number of non-overlapping occurrences of the respective events as M_{AC}, M_{ABC} and $M_{A\bar{B}C}$ respectively. (The arguments for overlapping episodes are similar.) If the connection AC is active, then $M_{A\bar{B}C}$ will be much more than what one would expect under independence. So we can test the hypothesis

$$\begin{aligned} H_0 &: P_{A\bar{B}C} \leq P_A P_{\bar{B}} P_C \\ H_1 &: P_{A\bar{B}C} > P_A P_{\bar{B}} P_C \end{aligned} \tag{11}$$

to determine if the AC connection is active or not. Let $\xi = P_{A\bar{B}C} - P_A P_{\bar{B}} P_C$; then the above

test is equivalent to testing whether $\xi \leq 0$ or $\xi > 0$.

Consider now the second type of false edge shown in the center panel of Figure 2. Again suppose that we have detected all three of the episodes $A[k_1]B$, $A[k_1 + k_2]C$, and $B[k_2]C$ as significant. Our goal is to test if the third episode is a false edge due to co-occurrences of the first two episodes (i.e., a single firing of A followed by both a firing of B and a firing of C with the specified delays).[1] We now use the difference between $P_{\bar{A}BC}$, the probability of the episode $B[k_2]C$ in the absence of A, and the probability of the same episode under independence, namely $P_{\bar{A}}P_BP_C$. The hypothesis test to eliminate this type of false edge is as follows:

$$\begin{aligned} H_0 &: \eta = P_{\bar{A}BC} - P_{\bar{A}}P_BP_C \leq 0 \\ H_1 &: \eta > 0 \end{aligned} \tag{12}$$

Denote the test statistics for these procedures as $Z_\xi = \hat{\xi}/\sqrt{\widehat{\mathrm{Var}}(\hat{\xi})}$ and $Z_\eta = \hat{\eta}/\sqrt{\widehat{\mathrm{Var}}(\hat{\eta})}$. The variances of these estimators are developed in Appendices A.2 and A.4. As before, we use a normal approximation of the standardized statistic to implement the procedure. One can develop other procedures for screening out false edges, but these tests are simple to implement when dealing with statistics based on non-overlapping occurrences, and we found them to perform well in simulation studies.

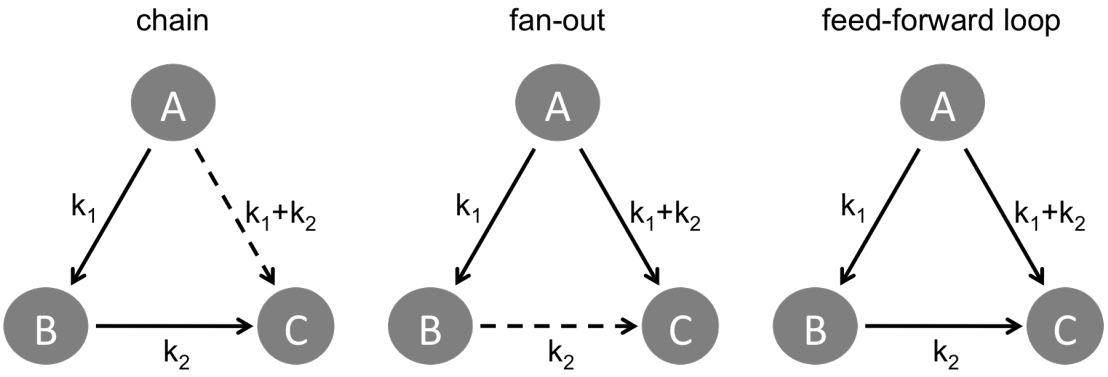

Figure 2: The three network structures that are likely to lead to frequent occurrences of the two-node episodes $A[k_1]B$, $B[k_2]C$, and $A[k_1 + k_2]C$. For chain and fan-out structures, one of these frequent episodes represents a false edge (dashed lines).

[1] The same arguments apply to detecting the first episode as a false edge, however in such a scenario there is no reason to expect the last two episodes to co-occur frequently.

5 Simulation Studies

To illustrate the usefulness of our results, we simulated data from a simple network of 9 neurons with a baseline firing rate of 5 Hz and the 7 connections shown in Figure 3. We varied the strength of the connections by setting S at values ranging from 1 to 40. Note that $S = 1$ corresponds to no connectivity as it implies that the neurons always spike according to the baseline firing rate.

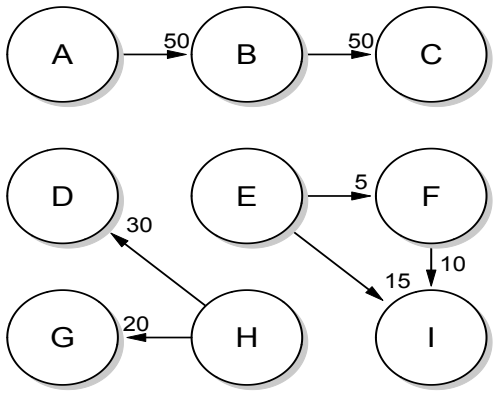

Figure 3: A network with 9 neurons and embedded connections. Each neuron had an intrinsic firing rate of 5 Hz. The connections had the time delays indicated next to each arrow (in ms), and strengths ranging from $S = 1$ to 40.

For each value of S, we simulated 100 data sets from the network. To analyze the data, we first count the total number of spikes from the 9 neurons; these can be viewed as 1-node episodes. We then obtain the count of all possible 2-node serial episodes with delays ranging from 1 to 200 ms. There are 81 pairwise episodes, so this is a total of $81 \times 200 = 16,200$ cases. We can then test for active 2-node connections. Since we are performing $n = 81 \times 200 = 16,200$ tests, we must consider the issue of multiple testing as described in Section 4.2.

We implemented our 2-node detection tests as described in Section 4. More specifically, we used the procedure in Equation (9) with different values of S_0 and examined the number of significant 2-node connections. The mean number of such connections for the 100 replicates are shown in Table 3. Consider first the cell corresponding to $S = S_0 = 1$. This is the null case where there are no connections, and we are testing the hypothesis of independence. At significance level $\alpha = 0.05$, we would expect about $16,200 \times 0.05$ or approximately 800

of these tests to be declared as significant. However, only 200 connections were identified as significant. One of the main reasons for this is that the $16,200$ tests are highly-positively dependent. In such cases, one would expect fewer number of tests to be identified as significant. For the row with $S_0 = 2$, we see that, on the average, 1 edge is falsely detected as significant; when $S_0 = 3$, we detect zero connections which is the correct value. Looking at moderate sized values of S (ranging from 2 to 4), we can see that the tests with $S_0 = 2$ miss some of the 7 active edges. When the connection strength is stronger ($S > 5$), the tests with $S_0 = 2$ detect at least 7 connections. A detailed analysis showed that all the active edges were indeed identified. For high values of S, we also detect some false edges. An inspection of Figure 3 shows that we may detect two false edges ($A[100]C$ and $G[10]D$) for the reasons discussed in Section 4.3. An occurrence of $A[100]C$ will happen whenever an occurrence of $A[50]B$ is followed by an occurrence of $B[50]C$. Similarly, co-occurrence of $H[20]G$ and $H[30]D$ will lead to an occurrence of $G[10]D$.

Table 3: Average number of significant 2-node connections for different connection strengths S and strength thresholds S_0 (network as shown in Figure 3, 100 replicates, $\alpha = 0.05$).

Connection	Strength threshold								
strength	$S_0 = 1$	2	3	4	5	10	15	20	25
$S = 1$	198.51	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	208.16	1.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	209.75	3.27	0.25	0.00	0.00	0.00	0.00	0.00	0.00
4	217.25	5.07	1.42	0.33	0.00	0.00	0.00	0.00	0.00
5	230.42	7.34	3.44	1.11	0.44	0.00	0.00	0.00	0.00
10	239.33	8.25	7.00	6.88	6.63	0.38	0.00	0.00	0.00
15	241.39	7.67	7.00	7.00	7.00	6.25	0.28	0.00	0.00
20	245.27	7.61	7.00	7.00	7.00	7.00	5.38	0.35	0.00
30	249.71	9.76	8.86	7.14	7.14	7.11	7.02	7.00	4.91
40	265.47	9.14	9.00	8.57	8.57	7.45	7.13	7.00	7.00

To demonstrate this, we discuss the case with $S = 30$ for one particular simulated data set. We applied all the test procedures, including the false-edge detection tests in Section 4.3. Table 4 shows the 9 connections that were significant based on their Z_τ values ($S_0 = 2$). Consider first the group of neurons A, B and C. We see that the episode $A[100]B$ has a negative Z_ξ value, indicating that it is a false edge, while $B[50]C$ has a large Z_η value, indicating that it is not a

false edge. Thus, we conclude these three neurons have a chain structure. Consider next the group of neurons D, H and G. Here, the large value of Z_ξ and negative value of Z_η for $H[30]D$ and $G[10]D$ respectively indicate that $G[10]D$ is a false edge while $H[30]D$ is not. Thus, we conclude these three neurons have a fan-out structure. Finally, consider the group of neurons E, F and I. Both the Z_ξ and Z_η values are large, indicating that neither $E[15]I$ nor $F[10]I$ are false edges. Thus, we conclude these three neurons have a feed-forward loop structure.

Table 4: Test statistics for the initial detection test (Z_τ) and the false edge elimination tests (Z_ξ, Z_η) for one replicate with $S = 30$.

Neuron1[k]Neuron2	Z_τ	Z_ξ	Z_η
A[50]B	15.76	-	-
A[100]C	7.85	-3.70	-
B[50]C	18.04	-	15.72
E[5]F	16.15	-	-
E[15]I	17.85	13.09	-
F[10]I	20.37	-	16.34
G[10]D	8.70	-	-1.38
H[30]D	16.36	13.13	-
H[20]G	17.32	-	-

5.1 Scalability

To assess the scalability of our approach, we analyzed simulated networks of various sizes and connection densities. We measured the accuracy and computational cost of our method, considering networks where the number of neurons ranged from $n = 10$ to 100, with $e = pn^2$ edges for $p = 0.01$ to 0.05. The connectivity was random, with connection strength $S = 30$ and delays of either 5 or 10 ms.

Accuracy

Table 5 shows the results of our methodology on data from networks of different sizes for a fixed density of connections ($p = 0.01$). With $S_0 = 2$ and considering delays from $k = 1$ to 20 ms, there were no false negatives for any n. However, as n increases, the number of false positives increases. The growth in false positives is substantially reduced by our false edge

elimination tests. For example, for the networks with $n = e = 100$, these additional hypothesis tests eliminated approximately 90% of the false positives.

We then explored how the positive predictive value (defined as the proportion of positive test results that are true positives) of our two-phase procedure scales with network size and connection density. If only the screening phase is used, then increasing network size and connection density p leads to a decrease in positive predictive value (PPV). However, when the false edge elimination phase is employed, PPV remains high for all the networks considered (see Figure 4). These results indicate that the performance of our two-phase methodology is relatively insensitive to increases in network size and connection density.

Table 5: Networks of n neurons with $e = 0.01n^2$ edges ($S = 30$, delays 5 or 10 ms) were simulated and analyzed with $S_0 = 2$ and $k = 1$ to 20 ms. For each configuration, 10 random networks were generated. The mean number of false positives (FP) is much lower when the false edge elimination tests are applied (pruned) than if they are not applied (unpruned).

Neurons	$n = 10$	20	30	40	50	60	70	80	90	100
Edges	$e = 1$	4	9	16	25	36	49	64	81	100
FP (unpruned)	0.2	1.3	3.5	8.2	18.4	25.8	42.6	64.4	86.2	119.0
FP (pruned)	0.2	0.7	1.0	1.9	4.4	4.0	6.1	6.7	8.0	12.1

Computational Cost

The bulk of the computational time involves counting all the two-node episodes in phase 1 of the strategy. We cannot think of any statistical technique for this problem where these counts will not be needed. Furthermore, as we have noted earlier, computationally-efficient algorithms have been developed for counting the non-overlapping occurrences of episodes [20, 21, 27]. Implementing the pruning step in phase 2 takes a relatively smaller amount of time. For example, a complete analysis of the largest and most dense network in Figure 4 ($n = 100, e = 500$) took 68.8 minutes on a standard laptop computer (MacBook Pro, 2.4 GHz Intel Core i7 processor, 8 GB 1600 MHz DDR3 memory). For networks with $p = 0.01$, counting the 3-node episodes required for implementation of phase 2 (false edge elimination tests) accounted for less than 1% of this computational time. These figures can be reduced on faster computational platforms and using parallel computing.

As the number of neurons n increases, the computational time scales linearly with the num-

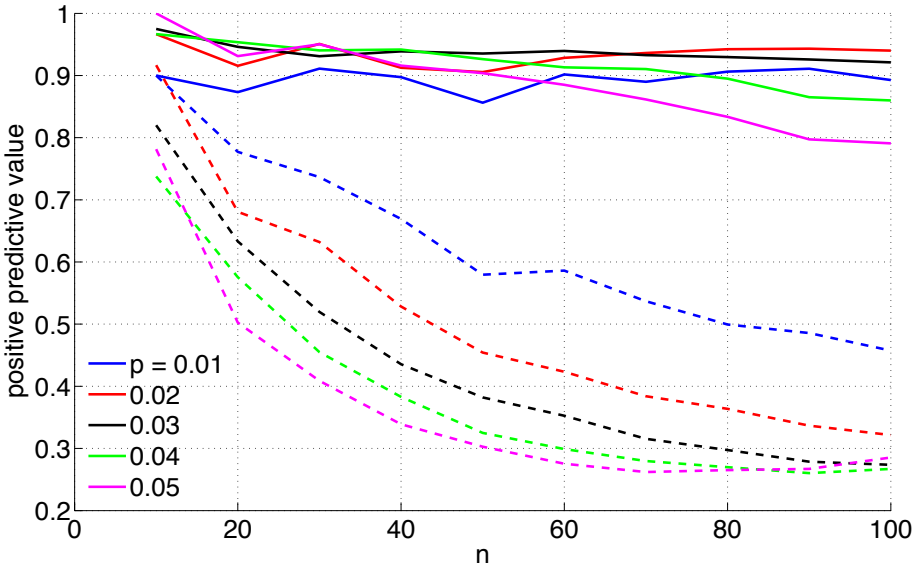

Figure 4: Scaling of PPV, defined as true positives divided by the sum of true positive and false positives, as a function of network size n and connection density p. Mean PPV for 10 random networks, with and without applying the false edge elimination tests (indicated by solid and dashed lines, respectively).

ber of 2-node episodes ($\sim n^2$) to be counted (see Figure 5). For networks with $n = 100$, as p is increased the amount of time spent counting 3-node episodes increases, but remains less than 10% of total computational time at $p = 0.05$.

Networks with non-DAG structure

We have assumed throughout that the underlying networks have a DAG structure. It is much more difficult to develop the distribution theory for the statistical procedures for non-DAG structures. We can do this under a specified non-DAG network structure, but it is not possible to do this under a general unspecified structure. Additionally, and more importantly, there can be identifiability as well as interpretability problems with non-DAG structures. For example, suppose $A \to B$, $B \to C$, and $C \to A$. Further, suppose that the first two connections are excitatory and the last one is inhibitory. Then, it becomes very difficult to untangle the opposing relationships and reconstruct the connections purely from the spike train data. Nevertheless, in reality one has to deal with non-DAG networks. So we have examined the potential limitation of this assumption through a simulation study. The simulation studies in Table 5 and Figure 4 were done under the assumption of a DAG network, i.e., the random connections were constrained to avoid

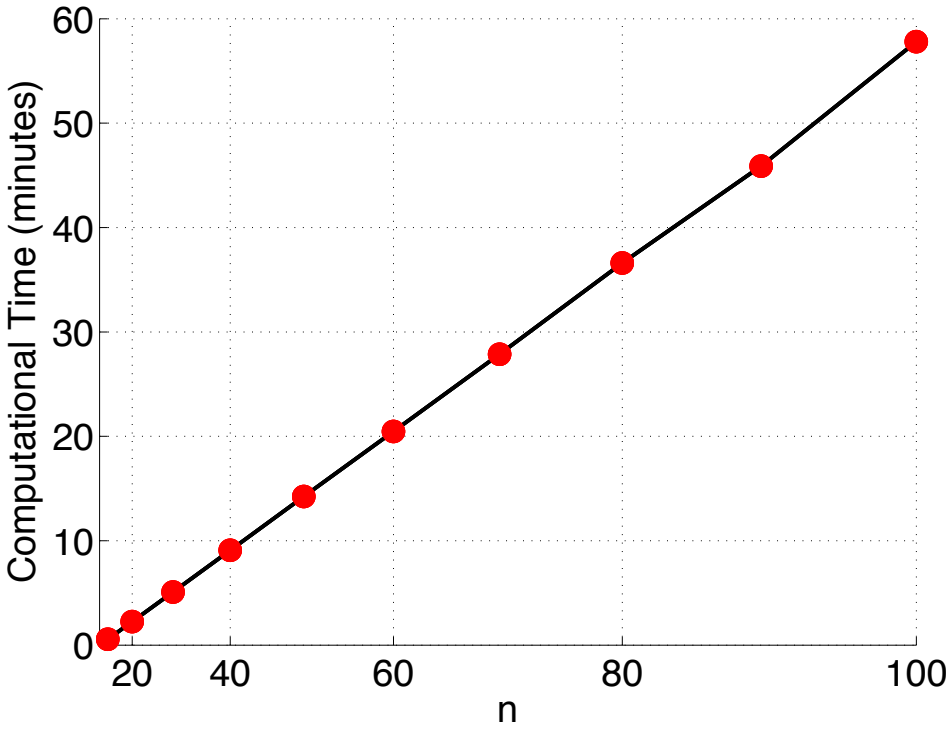

Figure 5: Computational time for analysis of the $p = 0.01$ networks shown in Figure 4. The time required scales linearly with the number of 2-node episodes to be counted, which is given by n^2 times the number of delays k being considered (here 20 different delays were used).

cycles. We repeated our analysis on non-DAG networks generated without this constraint. Figure 6 shows that the results for DAG and non-DAG networks are very similar in terms of PPV. While it is not possible to generalize from a single study, the results for these configurations are encouraging. However, as we have noted, this problem requires a more thorough investigation, both in terms of the performance on non-DAG structures and also the implications for identifiability and interpretability. We note that this is an issue not just for the methods in this paper, but for many others in the literature as well.

5.2 Effect of Subsampling

Spike train datasets recorded from mammalian tissues will typically be from a small subset of the neuronal network active in the tissue. This subsampling is due to the large number of neurons present in mammalian brain structures and technical limitations on the number of neurons that can be recorded from simultaneously. So it is possible (and even very likely) that the connections discovered in the observed network are not true connections and can be explained by chains and

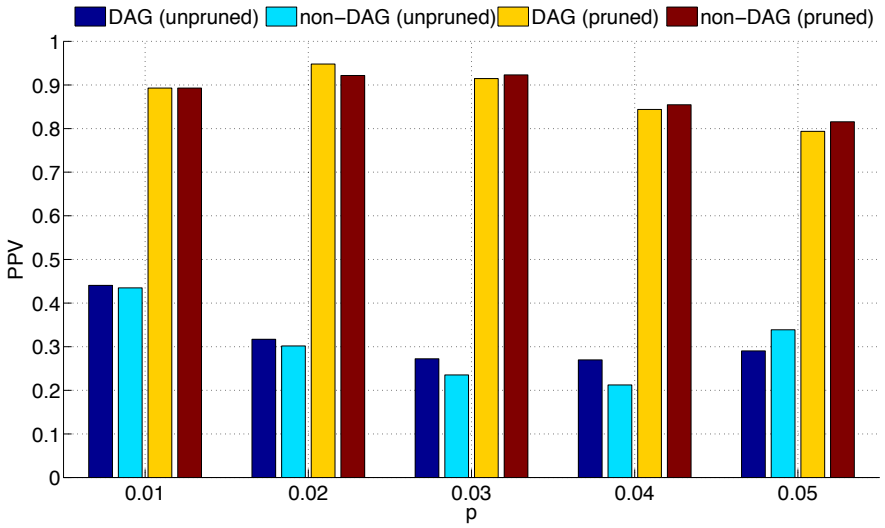

Figure 6: PPV for networks with $n = 100$ and various connection densities p is similar whether the random connections are unconstrained (non-DAG) or constrained to avoid cycles (DAG).

fan-out structures in the unobserved network. This problem is of course present in any statistical method that relies on data from the observed network. Thus, connections discovered in any observed network must be considered as "functional" connections and cannot be assumed to be a direct connection in the large network.

Nevertheless, we used a simulation study to assess the effect of unobserved neurons on the ability to discover functional connectivity. We simulated a network of 500 neurons and then randomly subsampled 20 of these neurons on which to perform our analysis. The original network had 2500 random connections of strength $S = 30$ and random delays of either 5 or 10 ms. For networks with 500 neurons and 2500 connections, we would expect 20-neuron subsamples to yield on average $(20^2/500^2)(2500) = 4$ direct connections between neurons in the subsample. For the network and subsample analyzed here, 3 of the 2500 connections were between neurons in our observed subsample. When we subject these 20 neurons to our analysis procedure, there are 9 significant 2-node episodes with $S_0 = 2$ (before false edge elimination). Three of these are true direct connections between neurons in the subsample: 5[10]11, 5[10]15, and 11[5]16. The 5[10]11 and 11[5]6 connections form a chain that lead to 5[15]6 being detected as significant; this is a false edge which is eliminated by hypothesis test (11). The other 5 episodes are due to chains or fan-outs involving unobserved neurons. For example, the significant episode 1[10]19 results from true edges 1[5]479 and 479[5]19, whereas the significant episode 13[5]4 comes from edges 186[5]13 and 186[10]14.

As the number of neurons sampled is increased from 20 to 100, the number of true positives (observed direct connections) increases, but so does the number of false positives. The majority of these false positives can be explained by the existence of chains where neuron B is unobserved, or fan-outs where neuron A is unobserved (see Figure 2 and Table 6). The estimated strengths of the false connections induced by the partially observed chains and fan-outs is much less than the estimated strengths of the true connections, thus moderately increasing S_0 is an effective means of reducing false positives while still avoiding false negatives.

Table 6: Subsamples of size $n_s = 20$ to 100 neurons from a 500-neuron network simulation (2500 edges, $S = 30$, delays 5 or 10 ms) were analyzed with $S_0 = 2$ to 4. Increasing S_0 did not affect the number of true positives (TP) or false negatives (FN), but did reduce the number of false positives. Most of these false positives are due to partial observation of chains and fan-out structures present in the 500-neuron network, i.e. cases where the middle neuron in a chain or the first neuron in a fan-out are not in the observed sample.

Subsamples	$n_s = 20$	30	40	50	60	70	80	90	100
TP	3	12	17	29	38	45	60	81	92
FN	0	0	0	0	0	0	0	0	0
FP ($S_0 = 2$)	5	13	27	45	71	95	116	144	163
FP ($S_0 = 3$)	0	1	4	5	11	18	24	30	34
FP ($S_0 = 4$)	0	1	2	2	4	4	4	6	7

6 Application to cultured cortical neurons

To understand the interactions of neurons in a network, neurobiologists conduct experiments in which they record the activity of several neurons from a region of the brain simultaneously. For *in vitro* experiments, the neurons often are often cultured on top of a grid of micro-electrodes known as a multi-electrode array (MEA). A typical MEA setup consists of 8×8 grid of 64 electrodes. Each electrode is capable of recording the activity of one or more neurons around it. Some of the brain areas of interest for such experiments are the primary motor cortex, hippocampus and pre-frontal cortex, as well as retinal ganglion cells.

After suitable pre-processing of the data, an MEA recording provides spike trains on each electrode which can be expressed as an event stream. We used our methods to analyze spike train data available in the literature on cultures of cortical neurons [32, 40]. In [40], "population bursts" in the spiking activity of the cultures were characterized. Such bursts, defined as brief

periods of time during which the firing rate of several cells or electrodes greatly exceeds the baseline rate, are a common feature observed in cultures of many different types of neurons. Our focus is not to characterize the bursts but rather to detect precisely timed spiking patterns involving multiple neurons in order to estimate the strength of functional connectivity between different neurons in the culture. Since our methods assume that the firing rates of individual neurons are relatively stationary in the analysis window, we selected a segment of the data that did not contain bursts.

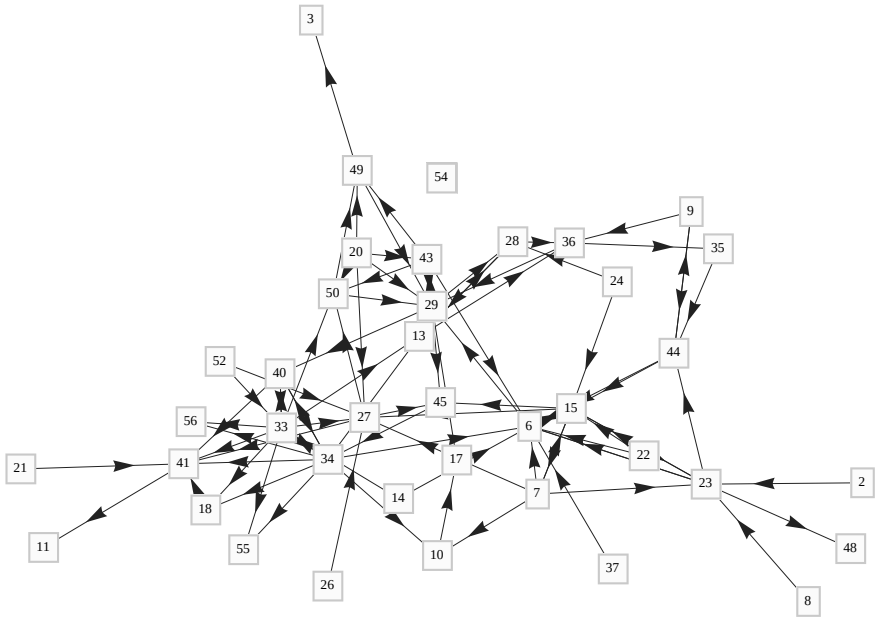

Figure 7: Inferred functional connectivity in cortical culture 2-1-34 with $S_0 = 2$.

We analyzed 120 seconds of data from culture 2-1-34. To visualize the functional connectivity present in the culture, we show the network graphs based on our pruned list of significant 2-node episodes. If there is a significant episode with any delay for a pair of neurons (i, j), we draw a directed edge $i \to j$ connecting them. Figures 7 and 8 show the inferred network structure for two different S_0 thresholds. As S_0 is increased from $S_0 = 2$ to $S_0 = 10$ (Figures 7 and 8), the weaker connections are no longer significant and the inferred network becomes more sparse. In this way, our methods can be used to identify the strongest microcircuits present in the culture. Although we considered episodes with delays of up to 200 ms, all the significant episodes had delays of less than 10 ms. Fast delays such as this are consistent with the timescale of the action for AMPA, a common excitatory neurotransmitter in the cortex [10, 29].

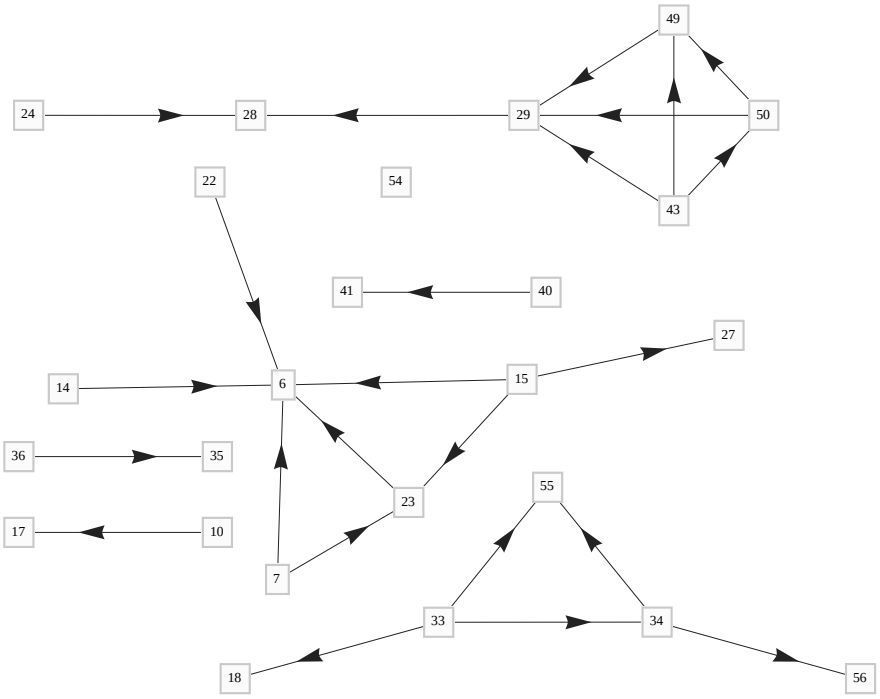

Figure 8: Inferred functional connectivity in cortical culture 2-1-34 with $S_0 = 10$.

7 Discussion

We have presented a two-phase approach for discovering functional connectivity of neuronal networks from spike trains based on repeating occurrences of temporally precise patterns of spikes. Most methods for detecting repeated occurrences of precise spike patterns and assessing their significance involve correlations between time-shifted spike trains [1, 37, 13, 30, 9]. Approaches utilizing dynamic Bayesian networks [28], Granger causality [19, 17, 18], and information theory [26, 14] have also been employed. In contrast, the framework developed here is based on a computationally-efficient temporal data mining method. Our hypothesis tests allow one to infer episodes that represent strong interactions among neurons, and also avoid certain systematic errors commonly made by network inference procedures [23]. We have demonstrated the effectiveness of our methods both on simulated and *in vitro* neuronal spike data. We do note that the methodology developed here can be characterized in terms of (simple) Granger causality methods.

The strength of connection between a pair of neurons is characterized using the conditional probability that one neuron fires after a specified delay given that the other neuron has fired. This

delay is usually caused by the axonal delay of spike propagation and the delay at synapses due to chemical diffusion. In our analysis, we have assumed that the inter-event time constraints which represent the time delays in neuronal connections are constants. In general there would be variation in these delays. Also, since we do not have complete control over which specific neurons are recorded from, we may be seeing some connections mediated by more than one synapse. Hence, one very useful extension for the method proposed here is to incorporate variable delays as inter-event time constraints.

Currently, our statistical framework handles only serial episodes. Another interesting problem would be to extend the statistical theory for the case of more general episodes which may be able to represent additional types of connectivity graphs. For example, [39] has developed a framework for determining the statistical significance of parallel episodes. Parallel episodes can capture the co-occurrence of spikes from many different neurons within a time-window, a type of synchronous activity of interest to many researchers [13], [30].

Acknowledgments

We are grateful to the reviewers and editors for their comments that have led to significant improvements. We thank Steve Potter for providing the culture data and Debprakash Patnaik for coding the counting algorithms used in this paper. K. P. U thanks P. S. Sastry for many useful discussions. C.O.D.'s work was supported in part by the Mathematical Biosciences Institute and the National Science Foundation under grant DMS 0931642. K.D's work was partially supported through a project funded by General Motors R&D Center, Warren, MI.

Appendix

Appendix A.1: Variance of $\tau_N(S_0)$ based on N

We derive the variance of the test statistic $\tau_N(S_0)$ based on the total number of occurrences. For simplicity of notation we will denote $A[k]B$ as AB. Recall that N_{AB} is binomial($L-k$, P_{AB}), where P_{AB} is the probability of an occurrence of the episode AB as defined in Section 3.1. Equation (10) for the variance is repeated below and it is a sum of three expressions

$$\mathrm{Var}(\hat{\tau}_N(S_0)) = \mathrm{Var}(\hat{P}_{AB}) + S_0^2 \,\mathrm{Var}(\hat{P}_A\hat{P}_B) - 2S_0 \,\mathrm{Cov}(\hat{P}_{AB}, \hat{P}_A\hat{P}_B). \tag{13}$$

From the binomial property of N_{AB}, the first term is

$$\mathrm{Var}(\hat{P}_{AB}) \;=\; \frac{P_{AB}(1-P_{AB})}{(L-k)}. \tag{14}$$

We will now show that the second term, up to the constant S_0^2 is

$$\begin{aligned}
\mathrm{Var}(\hat{P}_A\hat{P}_B) &= \frac{P_{AB}}{(L-k)^3}(1-P_{AB}-P_A-P_B+5P_AP_B) \\
&+ \frac{P_{AB}}{(L-k)^2}(P_{AB}+P_A+P_B-8P_AP_B) \\
&- \frac{P_AP_B}{(L-k)^3}(1-2P_A-2P_B+6P_AP_B) \\
&+ \frac{P_AP_B}{(L-k)^2}(1-3P_A-3P_B+10P_AP_B) \\
&+ \frac{P_AP_B}{(L-k)}(P_A+P_B+2P_{AB}-4P_AP_B), \qquad (15)
\end{aligned}$$

and the third term, up to the constant $-2S_0$ is

$$\begin{aligned}
\mathrm{Cov}(\hat{P}_{AB},\hat{P}_A\hat{P}_B) &= \frac{1}{(L-k)^2}(1-P_{AB}-P_{AB}P_A-P_{AB}P_B+2P_{AB}P_AP_B) \\
&- \frac{1}{(L-k)}(P_{AB}^2-P_{AB}P_A-P_{AB}P_B+2P_{AB}P_AP_B) \\
&+ P_{AB}^2. \qquad (16)
\end{aligned}$$

We start with the second term or Equation (15). Note that we are computing the variance of $\mathrm{Var}(\hat{P}_A\hat{P}_B)$ under possible dependence, so these two terms are not independent. We can write

$$\begin{aligned}
N_AN_B &= \sum_{t=1}^{L-k} I_A(t)\sum_{r=1}^{L-k} I_B(r+k) \\
&= \sum_{t=1}^{L-k} I_A(t)I_B(t+k) + \sum_{s=1}^{L-k}\sum_{r\neq s} I_A(s)I_B(r+k) \\
&= N_{AB}+N'_{AB},
\end{aligned}$$

Thus,

$$\mathrm{Var}(N_AN_B) = \mathrm{Var}(N_{AB}) + \mathrm{Var}(N'_{AB}) + 2\,\mathrm{Cov}(N_{AB},N'_{AB}).$$

Recall that $\mathrm{Var}(N_{AB}) = (L-k)P_{AB}(1-P_{AB})$. Further, it can be shown, through some tedious calculations, that $\mathrm{Cov}(N_{AB},N'_{AB}) = (L-k)(L-k-1)P_{AB}(P_A+P_B-2P_AP_B)$ and that

$$\begin{aligned}
\mathrm{Var}(N'_{AB}) &= (L-k)(L-k-1)P_AP_B(1-P_AP_B) \\
&+ (L-k)(L-k-1)(L-k-2)P_AP_B(P_A+P_B) \\
&+ (L-k)^2P_{AB}^2 + (L-k)(2(L-k)^2-6(L-k)+3)P_{AB}P_AP_B \\
&+ (L-k)(L-k-1)(L-k-2)(L-k-3)(P_AP_B)^2 \\
&- (L-k)(L-k-1)((L-k)^2-(L-k)-1)(P_AP_B)^2
\end{aligned}$$

We can now get Equation (15) by combining these expression and using the fact that $\mathrm{Var}(\hat{P}_A\hat{P}_B) = \frac{1}{(L-k)^4}\mathrm{Var}(N_AN_B)$.

Finally, consider the third term or Equation (16). Write

$$\mathrm{Cov}(N_{AB}, N_AN_B) = \mathrm{E}\,(N_{AB}N_AN_B) - \mathrm{E}\,(N_{AB})\mathrm{E}\,(N_AN_B). \tag{17}$$

We can express the first term on the right as

$$\begin{aligned}\mathrm{E}\,(N_{AB}N_AN_B) &= \mathrm{E}\,[\sum_{t=1}^{L-k} I_A(t)I_B(t+k)\sum_{s=1}^{L-k} I_A(s)I_B(s+k) \\ &+ \sum_{t=1}^{L-k} I_A(t)I_B(t+k)\sum_{s=1}^{L-k}\sum_{r\neq s} I_A(s)I_B(r+k)] \\ &= \mathrm{E}\,[N_{AB}^2] + \mathrm{E}\,[N_{AB}N'_{AB}] \end{aligned} \tag{18}$$

Further,

$$\mathrm{E}\,[N_{AB}^2] = (L-k)P_{AB}(1-P_{AB}) + (L-k)^2P_{AB}^2, \tag{19}$$

and

$$\begin{aligned}\mathrm{E}\,[N_{AB}N'_{AB}] &= \mathrm{E}\,[\sum_{t=1}^{L-k} I_A(t)I_B(t+k)\sum_{s=1}^{L-k}\sum_{r\neq s} I_A(s)I_B(r+k)] \\ &= (L-k)(L-k-1)P_{AB}(P_A+P_B+(L-k-2)P_AP_B)\end{aligned} \tag{20}$$

Moreover, the last term in Equation (17) (again noting that the two terms are not independent) can be calculated as

$$\begin{aligned}\mathrm{E}\,[N_AN_B] &= \mathrm{E}\,[N_{AB}] + \mathrm{E}\,[N'_{AB}] \\ &= (L-k)P_{AB} + (L-k)(L-k-1)P_AP_B.\end{aligned} \tag{21}$$

Combining the results in Equations (17)–(21) and noting that $\mathrm{Cov}(\hat{P}_{AB}, \hat{P}_A\hat{P}_B) = \mathrm{Cov}(N_{AB}, N_AN_B)/(L-k)^3$, we get Equation (16).

Appendix A.2: Variances of the test statistics for false edge elimination based on N

We first derive the variance of the test statistic $\hat{\xi}_N$ based on the total number of occurrences N. It can be written as the sum of three terms

$$\mathrm{Var}(\hat{\xi}_N) = \mathrm{Var}(\hat{P}_{A\bar{B}C}) + \mathrm{Var}(\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) - 2\,\mathrm{Cov}(\hat{P}_{A\bar{B}C}, \hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C). \tag{22}$$

Let k_1 and k_2 be the time delays, i.e., the event of interest is $A[k_1]\bar{B}[k_2]C$. From the binomial property of $N_{A\bar{B}C}$, the first term in Equation (22) is

$$\mathrm{Var}(\hat{P}_{A\bar{B}C}) = \frac{P_{A\bar{B}C}(1-P_{A\bar{B}C})}{(L-k_1-k_2)}. \tag{23}$$

We will show that the second term in Equation (22) is given by

$$\begin{aligned}
\mathrm{Var}(\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) &= \frac{1}{(L-k_1-k_2)^3}P_AP_{\bar{A}}P_BP_{\bar{B}}P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_{\bar{B}}^2P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_A^2P_BP_{\bar{B}}P_CP_{\bar{C}} \\
&+ \frac{1}{(L-k_1-k_2)}P_A^2P_{\bar{B}}^2P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)}P_A^2P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)}P_AP_{\bar{A}}P_{\bar{B}}^2P_C^2.
\end{aligned}$$

and that the third term Equation (22) is given by

$$\begin{aligned}
\mathrm{Cov}(\hat{P}_{A\bar{B}C},\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) &= \frac{1}{(L-k_1-k_2)^3}P_{A\bar{B}C}\Big[(1-P_{A\bar{B}C}-P_A-P_{\bar{B}}-P_C \\
&+ 2P_AP_{\bar{B}}+2P_AP_C+2P_{\bar{B}}P_C-6P_AP_{\bar{B}}P_C) \\
&+ \frac{1}{(L-k_1-k_2)^3}(P_{A\bar{B}C}+P_A+P_{\bar{B}}+P_C \\
&- 3P_AP_{\bar{B}}-3P_AP_C-3P_{\bar{B}}P_C) \\
&+ \frac{1}{(L-k_1-k_2)^2}(P_AP_{\bar{B}}+P_AP_C+P_{\bar{B}}P_C-6P_AP_{\bar{B}}P_C)\Big].
\end{aligned}$$

We start with the second term in Equation (22) and note that N_A,$N_{\bar{B}}$ and N_C are independent under H_0. Therefore,

$$\begin{aligned}
\mathrm{Var}(N_AN_{\bar{B}}N_C) &= \mathrm{Var}(N_A)\,\mathrm{Var}(N_{\bar{B}})\,\mathrm{Var}(N_C) + \mathrm{Var}(N_A)\,\mathrm{Var}(N_{\bar{B}})[\mathrm{E}\,(N_C)]^2 \\
&+ \mathrm{Var}(N_A)\,\mathrm{Var}(N_C)[\mathrm{E}\,(N_{\bar{B}})]^2 + \mathrm{Var}(N_{\bar{B}})\,\mathrm{Var}(N_C)[\mathrm{E}\,(N_A)]^2 \\
&+ [\mathrm{E}\,(N_A)]^2[\mathrm{E}\,(N_{\bar{B}})]^2\,\mathrm{Var}(N_C) + [\mathrm{E}\,(N_A)]^2[\mathrm{E}\,(N_C)]^2\,\mathrm{Var}(N_{\bar{B}}) \\
&+ [\mathrm{E}\,(N_{\bar{B}})]^2[\mathrm{E}\,(N_C)]^2\,\mathrm{Var}(N_A)
\end{aligned}$$

We can get the answer for the second term as

$$\mathrm{Var}(\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) = \frac{1}{(L-k_1-k_2)^6}\,\mathrm{Var}(N_AN_{\bar{B}}N_C) \tag{24}$$

Turning to the third term in Equation (22),

$$\begin{aligned}
\mathrm{Cov}(N_{A\bar{B}C},N_AN_{\bar{B}}N_C) &= \mathrm{E}\,(N_{A\bar{B}C}N_AN_{\bar{B}}N_C) - \mathrm{E}\,(N_{A\bar{B}C})\mathrm{E}\,(N_AN_{\bar{B}}N_C) \\
\mathrm{E}\,(N_{A\bar{B}C}N_AN_{\bar{B}}N_C) &= \mathrm{E}\,[N_{A\bar{B}C}^2] + \mathrm{E}\,[N_{A\bar{B}C}N'_{A\bar{B}C}]
\end{aligned} \tag{25}$$

These expectations can be obtained as

$$\mathrm{E}\left[N_{A\bar{B}C}^2\right] = (L-k_1-k_2)(1-P_{A\bar{B}C}) + (L-k_1-k_2)^2 P_{A\bar{B}C}^2 \tag{26}$$

and

$$\begin{aligned}
\mathrm{E}\left[N_{A\bar{B}C}N'_{A\bar{B}C}\right] &= \mathrm{E}\Big[\sum_{t=1}^{L-k_1-k_2} I_A(t)I_{\bar{B}}(t+k_1)I_C(t+k_1+k_2) \\
&\times \sum_{t=1}^{L-k_1-k_2}\sum_{s\neq t}\sum_{r\neq s\neq t} I_A(t)I_{\bar{B}}(s+k_1)I_C(r+k_1+k_2)\Big] \\
&= (L-k_1-k_2)(L-k_1-k_2-1)P_{A\bar{B}C}[P_A+P_{\bar{B}}+P_C \\
&+ (L-k_1-k_2-2)P_AP_{\bar{B}} \\
&+ (L-k_1-k_2-2)P_AP_C + (L-k_1-k_2-2)P_{\bar{B}}P_C \\
&+ (L-k_1-k_2-2)(L-k_1-k_2-3)P_AP_{\bar{B}}P_C].
\end{aligned} \tag{27}$$

Combining Equations (25)–(27) and using the fact that

$$\mathrm{Cov}(\hat{P}_{A\bar{B}C}, \hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) = \frac{\mathrm{Cov}(N_{A\bar{B}C}, N_AN_{\bar{B}}N_C)}{(L-k_1-k_2)^4}, \tag{28}$$

we get the final answer.

The variance of η_N can be obtained using similar arguments. We provide just the expressions and omit the derivations. Thus,

$$\mathrm{Var}(\hat{\eta}_N) = \mathrm{Var}(\hat{P}_{\bar{A}BC}) + \mathrm{Var}(\hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C) - 2\,\mathrm{Cov}(\hat{P}_{\bar{A}BC}, \hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C). \tag{29}$$

The first term in Equation (29) is

$$\mathrm{Var}(\hat{P}_{\bar{A}BC}) = \frac{P_{\bar{A}BC}(1-P_{\bar{A}BC})}{(L-k_1-k_2)} \tag{30}$$

The second term is

$$\begin{aligned}
\mathrm{Var}(\hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C) &= \frac{1}{(L-k_1-k_2)^3}P_AP_{\bar{A}}P_BP_{\bar{B}}P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_B^2P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_{\bar{A}}^2P_BP_{\bar{B}}P_CP_{\bar{C}} \\
&+ \frac{1}{(L-k_1-k_2)}P_{\bar{A}}^2P_B^2P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)}P_A^2P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)}P_AP_{\bar{A}}P_B^2P_C^2
\end{aligned}$$

Further,

$$\begin{aligned}\mathrm{Cov}(\hat{P}_{\bar{A}BC}, \hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C) &= \frac{1}{(L-k_1-k_2)^3} P_{\bar{A}BC}\Big[(1 - P_{\bar{A}BC} - P_{\bar{A}} - P_B - P_C \\ &+ 2P_{\bar{A}}P_B + 2P_{\bar{A}}P_C + 2P_BP_C - 6P_{\bar{A}}P_BP_C) \\ &+ \frac{1}{(L-k_1-k_2)^3}(P_{\bar{A}BC} + P_{\bar{A}} + P_B + P_C \\ &- 3P_{\bar{A}}P_B - 3P_{\bar{A}}P_C - 3P_BP_C) \\ &+ \frac{1}{(L-k_1-k_2)^2}(P_{\bar{A}}P_B + P_{\bar{A}}P_C + P_BP_C - 6P_AP_BP_C)\Big].\end{aligned}$$

Appendix A.3: Variance of $\tau_M(S_0)$ based on M

All the estimators in this section are based on the number of non-overlapping occurrences of an episode, denoted M_{AB}. We compute the variance of the test statistic $\hat{\tau}_M(S_0)$ which is the sum of three terms

$$\mathrm{Var}(\hat{\tau}_M) = \mathrm{Var}(\hat{P}_{AB}) + S_0^2\,\mathrm{Var}(\hat{P}_A\hat{P}_B) - 2S_0\,\mathrm{Cov}(\hat{P}_{AB}, \hat{P}_A\hat{P}_B), \tag{31}$$

where the first term is

$$\mathrm{Var}(\hat{P}_{AB}) \approx \frac{(1+kP_{AB})P_{AB}(1-P_{AB})}{(L-k)}, \tag{32}$$

the second term, up to the constant S_0^2, is

$$\begin{aligned}\mathrm{Var}(\hat{P}_A\hat{P}_B) &= \frac{P_B}{(L-k)^3}(1 - P_{AB} - P_A - P_B + 5P_AP_B) \\ &+ \frac{P_{AB}}{(L-k)^2}(P_{AB} + P_A + P_B - 8P_AP_B) \\ &- \frac{P_AP_B}{(L-k)^3}(1 - 2P_A - 2P_B + 6P_AP_B) \\ &+ \frac{P_AP_B}{(L-k)^2}(1 - 3P_A - 3P_B + 10P_AP_B) \\ &+ \frac{P_AP_B}{(L-k)}(P_A + P_B + 2P_{AB} - 4P_AP_B)\end{aligned} \tag{33}$$

and the third term, up to the constant $-2S_0$, is

$$\begin{aligned}\mathrm{Cov}(\hat{P}_{AB}, \hat{P}_A\hat{P}_B) &\approx (1+kP_{AB})\Big[\frac{P_{AB}(1-P_{AB})}{(L-k)^2(1+kP_{AB})} \\ &- \frac{1}{(L-k)^2}(P_{AB}P_A + P_{AB}P_B - 2P_AP_B) \\ &- \frac{1}{(L-k)}(P_{AB}^2 + P_{AB}P_A + P_{AB}P_B - 3P_AP_B) \\ &+ P_AP_B\Big]\end{aligned} \tag{34}$$

We now derive each of these expressions. From Equation (7), the first term is

$$\mathrm{Var}(\hat{P}_{AB}) \approx \frac{(1+kP_{AB})P_{AB}(1-P_{AB})}{(L-k)} \tag{35}$$

The second term, $\mathrm{Var}(\hat{P}_A\hat{P}_B)$, is the same as in Appendix A.1. To compute the third term, write

$$\begin{aligned}
\mathrm{Cov}(M_{AB}, N_AN_B) &= \mathrm{E}\,(M_{AB}N_AN_B) - \mathrm{E}\,(M_{AB})\mathrm{E}\,(N_AN_B) \\
\mathrm{E}\,(M_{AB}N_AN_B) &= \mathrm{E}\,[\sum_{t^*=1}^{L-k} I_A(t^*)I_B(t^*+k) \sum_{s=1}^{L-k} I_A(s)I_B(s+k) \\
&+ \sum_{t^*=1}^{L-k} I_A(t^*)I_B(t^*+k) \sum_{s=1}^{L-k}\sum_{r\neq s} I_A(s)I_B(r+k)] \\
&= \mathrm{E}\,[M_{AB}N_{AB}] + \mathrm{E}\,[M_{AB}N'_{AB}]
\end{aligned} \tag{36}$$

Here t^* refers to the non-overlapped t's, and there are M_{AB} of them. After some algebraic manipulations, we get

$$\begin{aligned}
\mathrm{Cov}(M_{AB}, N_AN_B) &= \frac{(L-k)P_{AB}}{(1+kP_{AB})}\Big[(L-k-1)(P_A+P_B+(L-k-2)P_AP_B) \\
&+ (1+kP_{AB}) + (1-P_{AB})/(1+kP_{AB}) \\
&- (L-k)P_{AB} + (L-k)(L-k-1)P_AP_B\Big]
\end{aligned}$$

Since

$$\mathrm{Cov}(\hat{P}_{AB}, \hat{P}_A\hat{P}_B) \approx \frac{\mathrm{Cov}(M_{AB}, N_AN_B)}{(L-k)\mathrm{E}\,[M_{AB}]^2(\frac{L-k}{\mathrm{E}\,[M_{AB}]}-k)^2}, \tag{37}$$

we get the final answer.

Appendix A.4: Variances of the test statistics for false edge elimination based on M

We derive the expressions for the test statistics based on the number of non-overlapping occurrences M. As before, the variance of $\hat{\xi}_M$ can be written as the sum of three terms

$$\mathrm{Var}(\hat{\xi}_M) = \mathrm{Var}(\hat{P}_{A\bar{B}C}) + \mathrm{Var}(\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) - 2\,\mathrm{Cov}(\hat{P}_{A\bar{B}C}, \hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C), \tag{38}$$

where the first term is

$$\mathrm{Var}(\hat{P}_{A\bar{B}C}) \approx \frac{(1+(k_1+k_2)P_{A\bar{B}C})P_{A\bar{B}C}(1-P_{A\bar{B}C})}{(L-k_1-k_2)}, \tag{39}$$

the second term is

$$\begin{aligned}
\mathrm{Var}(\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) &= \frac{1}{(L-k_1-k_2)^3}P_AP_{\bar{A}}P_BP_{\bar{B}}P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_{\bar{B}}^2P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_A^2P_BP_{\bar{B}}P_CP_{\bar{C}} \\
&+ \frac{1}{(L-k_1-k_2)}P_A^2P_{\bar{B}}^2P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)}P_A^2P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)}P_AP_{\bar{A}}P_{\bar{B}}^2P_C^2, \qquad (40)
\end{aligned}$$

and the third term is

$$\begin{aligned}
\mathrm{Cov}(\hat{P}_{A\bar{B}C},\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C) &\approx \frac{P_{A\bar{B}C}(1+(k_1+k_2)P_{A\bar{B}C})}{(L-k_1-k_2)^2}\Big[\frac{(1-P_{A\bar{B}C})}{1+k_1+k_2P_{A\bar{B}C}} \\
&- (P_A+P_{\bar{B}}+P_C-2P_AP_{\bar{B}}-2P_AP_C-2P_{\bar{B}}P_C+6P_AP_{\bar{B}}P_C) \\
&+ (L-k_1-k_2)(P_A+P_{\bar{B}}+P_C-3P_AP_{\bar{B}} \\
&- 3P_AP_C-3P_{\bar{B}}P_C+11P_AP_{\bar{B}}P_C) \\
&+ (L-k_1-k_2)^2(P_AP_{\bar{B}}+P_AP_C+P_{\bar{B}}P_C-6P_AP_{\bar{B}}P_C)\Big]. \qquad (41)
\end{aligned}$$

The expression for $\mathrm{Var}(\hat{P}_{A\bar{B}C})$ can be derived in exactly the same way as that of $\mathrm{Var}(\hat{P}_{AB})$ in Section 3.3. The expression for $\mathrm{Var}(\hat{P}_A\hat{P}_{\bar{B}}\hat{P}_C)$ was already obtained in Appendix A.2. Turning to the third term, note that

$$\mathrm{Cov}(M_{A\bar{B}C},N_AN_{\bar{B}}N_C) = \mathrm{E}\,(M_{A\bar{B}C}N_AN_{\bar{B}}N_C) - \mathrm{E}\,(M_{A\bar{B}C})\mathrm{E}\,(N_AN_{\bar{B}}N_C),$$

where

$$\mathrm{E}\,(M_{A\bar{B}C}N_AN_{\bar{B}}N_C) = \mathrm{E}\,[M_{A\bar{B}C}N_{A\bar{B}C}] + \mathrm{E}\,[M_{A\bar{B}C}N'_{A\bar{B}C}].$$

Now,

$$\begin{aligned}
\mathrm{E}\,[M_{A\bar{B}C}N_{A\bar{B}C}] &= \mathrm{E}\,[M_{A\bar{B}C}(M_{A\bar{B}C}+\sum_{j=1}^{M_{A\bar{B}C}}R_j)] \\
&= \mathrm{E}\,[M_{A\bar{B}C}^2] + \mathrm{E}\,[M_{A\bar{B}C}\sum_{j=1}^{M_{A\bar{B}C}}R_j] \\
&= \mathrm{E}\,[M_{A\bar{B}C}^2(1+kP_{AB})], \qquad (42)
\end{aligned}$$

and

$$
\begin{aligned}
E[M_{A\bar{B}C}N'_{A\bar{B}C}] &= E[\sum_{t^*=1}^{L-k_1-k_2} I_A(t^*)I_{\bar{B}}(t^*+k_1)I_C(t^*+k_1+k_2) \\
&\times \sum_{t=1}^{L-k_1-k_2}\sum_{s\neq t}\sum_{r\neq s\neq t} I_A(t)I_{\bar{B}}(s+k_1)I_C(r+k_1+k_2)] \qquad (43)\\
&= \frac{(L-k_1-k_2)P_{A\bar{B}C}}{(1+(k_1+k_2)P_{A\bar{B}C})}[(L-k_1-k_2-1)(P_A+P_{\bar{B}}+P_C) \\
&+ (L-k_1-k_2-1)(L-k_1-k_2-2)(P_AP_{\bar{B}}+P_AP_C+P_{\bar{B}}P_C) \\
&+ (L-k_1-k_2-1)(L-k_1-k_2-2)(L-k_1-k_2-3)(P_AP_{\bar{B}}P_C)].
\end{aligned}
$$

Here t^* refers to the non-overlapped t's, and there are $M_{A\bar{B}C}$ of them. We get the final result by combining these equations.

The variance of $\hat{\eta}_M$ in Equation (12) can be obtained similarly. We omit the details and provide only the expressions.

$$
\mathrm{Var}(\hat{\eta}_M) = \mathrm{Var}(\hat{P}_{\bar{A}BC}) + \mathrm{Var}(\hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C) - 2\,\mathrm{Cov}(\hat{P}_{\bar{A}BC}, \hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C).
$$

where

$$
\begin{aligned}
\mathrm{Var}(\hat{P}_{\bar{A}BC}) &\approx \frac{(1+k_1+k_2P_{\bar{A}BC})P_{\bar{A}BC}(1-P_{\bar{A}BC})}{(L-k_1-k_2)}, \\
\mathrm{Var}(\hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C) &= \frac{1}{(L-k_1-k_2)^3}P_AP_{\bar{A}}P_BP_{\bar{B}}P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)^2}P_AP_{\bar{A}}P_B^2P_CP_{\bar{C}} + \frac{1}{(L-k_1-k_2)^2}P_{\bar{A}}^2P_BP_{\bar{B}}P_CP_{\bar{C}} \\
&+ \frac{1}{(L-k_1-k_2)}P_{\bar{A}}^2P_B^2P_CP_{\bar{C}} + \frac{1}{(L-2k)}P_{\bar{A}}^2P_BP_{\bar{B}}P_C^2 \\
&+ \frac{1}{(L-k_1-k_2)}P_AP_{\bar{A}}P_B^2P_C^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(\hat{P}_{\bar{A}BC}, \hat{P}_{\bar{A}}\hat{P}_B\hat{P}_C) &\approx \frac{P_{\bar{A}BC}(1+(k_1+k_2)P_{\bar{A}BC})}{(L-k_1-k_2)^2}\Big[(1-P_{\bar{A}BC})/(1+(k_1+k_2)P_{\bar{A}BC}) \\
&- (P_{\bar{A}}+P_B+P_C-2P_AP_B-2P_{\bar{A}}P_C-2P_BP_C+6P_{\bar{A}}P_BP_C) \\
&+ (L-k_1-k_2)(P_{\bar{A}}+P_B+P_C-3P_{\bar{A}}P_B \\
&- 3P_{\bar{A}}P_C-3P_BP_C+11P_{\bar{A}}P_BP_C) \\
&+ (L-k_1-k_2)^2(P_{\bar{A}}P_B+P_{\bar{A}}P_C+P_BP_C-6P_{\bar{A}}P_BP_C)\Big].
\end{aligned}
$$

References

[1] M. Abeles and G.L. Gerstein. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *Journal of Neurophysiology*, 60(3):909–924, 1988.

[2] A. Achar, S. Laxman, and P. .S. Sastry. A unified view of the apriori-based algorithms for frequent episode discovery. *Knowl Inf Syst*, doi:10.1007/s10115-011-0408-2, 2011.

[3] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298:824–827, 2002.

[4] S. Amari, H. Nakahara, S. Wu, and Y. Sakai. Synchronous firing and higher-order interactions in neuron pool. *Neural Computation*, 15:127–142, 2003.

[5] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57:289-300, 1995.

[6] D.R. Brillinger. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59:189–200, 1988.

[7] E.M. Brown, R.E. Kass, and P.M. Mitra. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience*, 7(5):456–461, 2004.

[8] E. Chornoboy, L. Schramm, and A. Karr. Maximum likelihood identification of neuronal point process systems. *Biological Cybernetics*, 59:265–275, 1988.

[9] C.O. Diekman, P.S. Sastry, and K.P. Unnikrishnan. Statistical significance of sequential firing patterns in multi-neuronal spike trains. *J Neurosci Methods*, 182:279–284, 2009.

[10] A. Destexhe and T. Sejnowski. *Thalamocortical assemblies: How ion channels, single neurons, and large-scale networks organize sleep oscillations*. Oxford University Press, New York, 2001.

[11] A.L. Fleishman. A method for simulating non-normal distributions. *Psychometrika*, 43(4):521-532, 1978.

[12] F. Gerhard, G. Pipa, B. Lima, S.Neuenschwander, and W. Gerstner. Extraction of network topology from multi-electrode recordings: is there a small-world effect? *Frontiers in Computational Neuroscience*, 5:4, 2011.

[13] S. Grün, M. Diesmann, and A. Aertsen. Unitary events in multiple single-neuron spiking activity: I. Detection and significance. *Neural Computation*, 14(4):43-80, 2001.

[14] S. Ito, M. E. Hansen, R. Heiland, A. Lumsdaine, A. M. Litke, and J. M. Beggs. Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. *PLoS One*, 6(11): e27431, 2011.

[15] R.E. Kass and V. Ventura. A spike-train probability model. *Neural Computation*, 13:1713-1720, 2001.

[16] R.E. Kass, V. Ventura and E.N. Brown. Statistical issues in the analysis of neuronal data. *J Neurophysiol*, 94:8-25, 2005.

[17] S. Kim, D. Putrino, S. Ghosh, and E.N. Brown. A Granger causality measure for point process models of ensemble neural spiking activity. *PLOS Comput Biol*, 7:e1001110, 2011.

[18] T. Kispersky, G. J. Gutierrez, and E. Marder. Functional connectivity in a rhythmic inhibitory circuit using Granger causality. *Neural Systems & Circuits*, 1:9, 2011.

[19] M. Krumin and S. Shoham. Multivariate autoregressive modeling and granger causality analysis of multiple spike trains. *Comput Intell Neurosci*, doi:10.1155/2010/752428, 2010.

[20] S. Laxman, P.S. Sastry, and K.P. Unnikrishnan. Discovering frequent episodes and learning hidden markov models: A formal connection. In *IEEE Transactions on Knowledge and Engineering*, 17:1505–1517, 2005.

[21] S. Laxman, P.S. Sastry, and K.P. Unnikrishnan. A fast algorithm for finding frequent episodes in event streams. In *Data Mining and Knowledge Discovery*, 2007.

[22] H. Mannila, H. Toivonen, and A. Verkamo. Discovery of frequent episodes in event sequences. In *Data Mining and Knowledge Discovery*, 1:259–289, 1997.

[23] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitsky. Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. USA*, 107: 6286–6291, 2010.

[24] L. Martignon, G. Deco, K. Laskey, M. Diamond, W. Freiwald, and E. Vaadia. Neural coding: Higher-order temporal patterns in the neurostatistics of cell assemblies. *Neural Computation*, 12:2621-2653, 2000.

[25] C.R. Olson, S.N. Gettner, V. Ventura, R. Carta, and R.E. Kass. Neuronal activity in macaque supplementary eye field during planning of saccades in response to pattern and spatial cues. *J Neurophysiol*, 84:1369–1384, 2000.

[26] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *J Comput Neurosci*, 30:17–44, 2011.

[27] D. Patnaik, P.S. Sastry, and K.P. Unnikrishnan. Inferring neuronal network connectivity from spike data: A temporal datamining approach. *Scientific Programming*, 16(1):49–77, 2008.

[28] D. Patnaik, S. Laxman, and N. Ramakrishnan. Discovering excitatory relationships using dynamic Bayesian networks. *Knowl Inf Syst*, 29:273–303, 2011.

[29] S. R. Platt. The role of glutamate in central nervous system health and disease – A review. *Vet J*, 173:278-286, 2007.

[30] G. Pipa, D.W. Wheeler, W. Singer, D. Nikolic. NeuroXidence: reliable and efficient analysis of an excess or deficiency of joint spike events. *J Comput Neurosci*, 25:64–88, 2008.

[31] F. Rigat, M. Gunst, and J. Pelt. Bayesian modelling and analysis of spatio-temporal neuronal networks. *Bayesian Analysis*, 1:733–764, 2006.

[32] J. Rolston, D.A. Wagenaar, and S.M. Potter. Precisely timed spatiotemporal patterns of neural activity in dissociated cell cultures. *Neuroscience*, 148(1):294–303, 2007.

[33] P.S. Sastry and K.P. Unnikrishnan. Conditional probability-based significance tests for sequential patterns in multineuronal spike trains. *Neural Computation*, 22:1025–1059, 2010.

[34] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–1012, 2006.

[35] M.N. Shadlen and W.T. Newsome. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J Neurosci*, 18(10):3870–3896, 1998.

[36] J.P. Shaffer. Multiple hypothesis testing. *Ann Rev Psych*, 46:561–584, 1995.

[37] I.V. Tetko and A.E.P. Villa. A pattern grouping algorithm for analysis of spatiotemporal patterns in neuronal spike trains: 1. Detection of repeated patterns. *J Neurosci Methods*, 105:1–14, 2001.

[38] W. Truccolo, U. Eden, M. Fellows, J. Donoghue, and E. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *J Neurophysiol*, 93:1074–1089, 2005.

[39] R. Viswanathan, P. S. Sastry, and K. P. Unnikrishnan. Efficient discovery of large synchronous events in neural spike streams. *arXiv*: 1006.1543v1 [cs.NE], 2010.

[40] D.A. Wagenaar, J. Pine, and S.M. Potter. An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC Neurosci*, 7(11), 2006.

[41] Y.G. Zhao, T. Ono and K. Ishii. Monte Carlo simulation using moments of random variables. *Journal of Asian Architecture and Building Engineering*, March 2002.