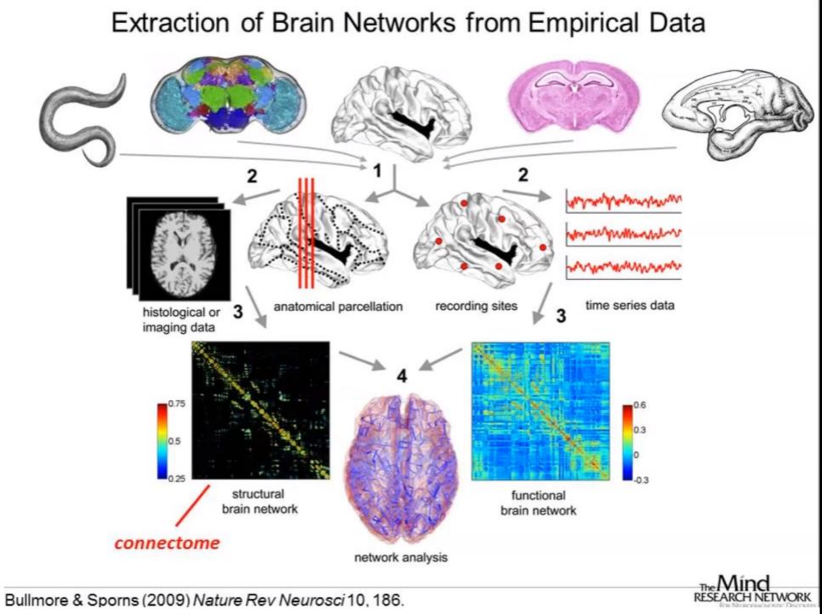
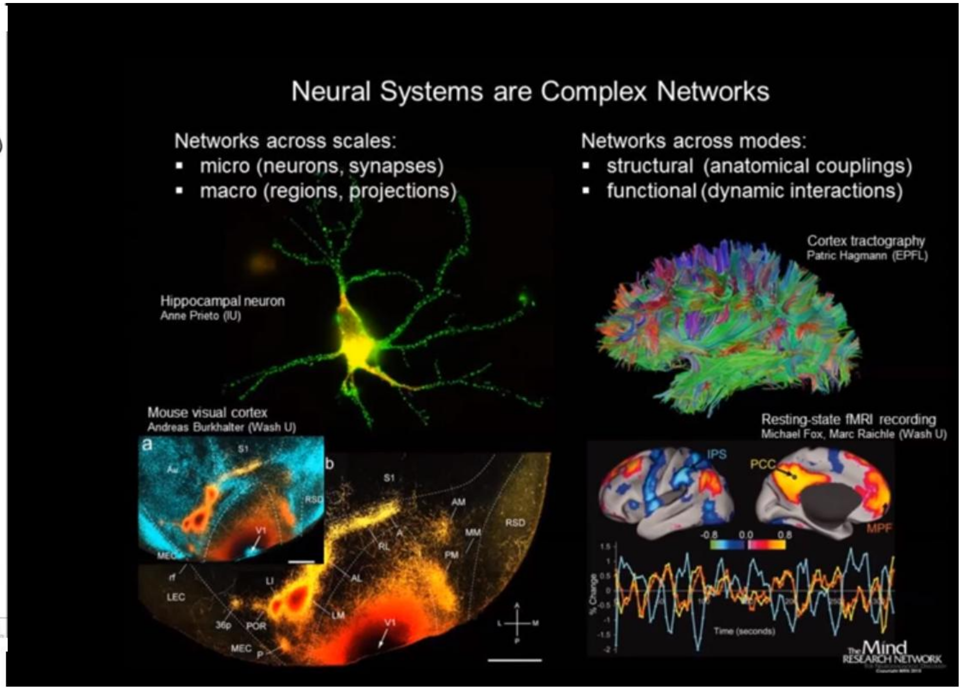




Lecture on Temporal Correlation, Kolmogorov-Smirnov Test & K-means

CS – 590.21 Analysis and Modeling of Brain Networks

Department of Computer Science
University of Crete

Greek Diaspora Fellowship Program

ΙΣΝ / SNF
ΙΔΡΥΜΑ ΣΤΑΥΡΟΣ ΝΙΑΡΧΟΣ
STAVROS NIARCHOS FOUNDATION


EVERYBODY WHO WENT TO THE MOON HAS EATEN CHICKEN!
GOOD GRIEF. CHICKEN MAKES YOU GO TO THE MOON!

Challenges in Quantifying Correlation

1. Correlated neurons fire at **similar times but not precisely synchronously**, so correlation must be defined with **reference to a timescale** within which spikes are considered correlated
2. Spiking is sparse with respect to the recording's sampling frequency & spike duration

   e.g., spiking rate 1 Hz, sampling rate typically 20 kHz (Demas et al., 2003)

This means that conventional approaches to correlation (such as Pearson's correlation coefficient) are unsuitable

- **as periods of quiescence should not count as correlated**
- correlations should **compare spike trains over short timescales, not just instantaneously.**

Pearson Correlation of two variables X & Y ($\rho_{X,Y}$)

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

where:

- cov is the covariance
- σ_X is the standard deviation of X
- σ_Y is the standard deviation of Y

ne formula for ρ can be expressed in terms of mean and expectation. Since

$$\mathrm{cov}(X,Y) = \mathbf{E}[(X-\mu_X)(Y-\mu_Y)],$$ [5]

e formula for ρ can also be written as

$$\rho_{X,Y} = \frac{\mathbf{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

where:

- cov and σ_X are defined as above
- μ_X is the mean of X
- $\mathbf{E}$ is the expectation.

Sample Pearson correlation coefficient

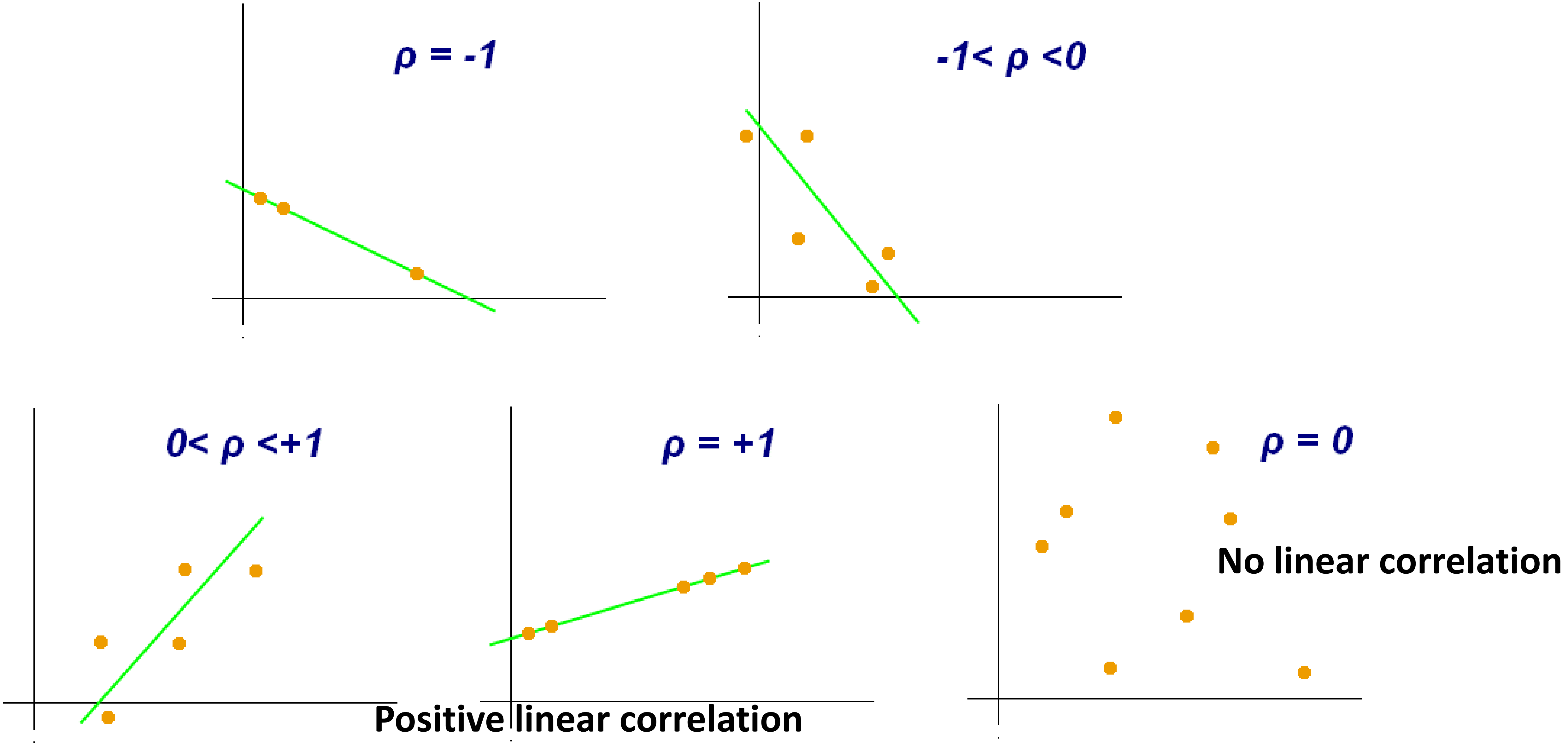
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Datasets $\{x_1,\ldots,x_n\}$ & $\{y_1,\ldots,y_n\}$ containing n values

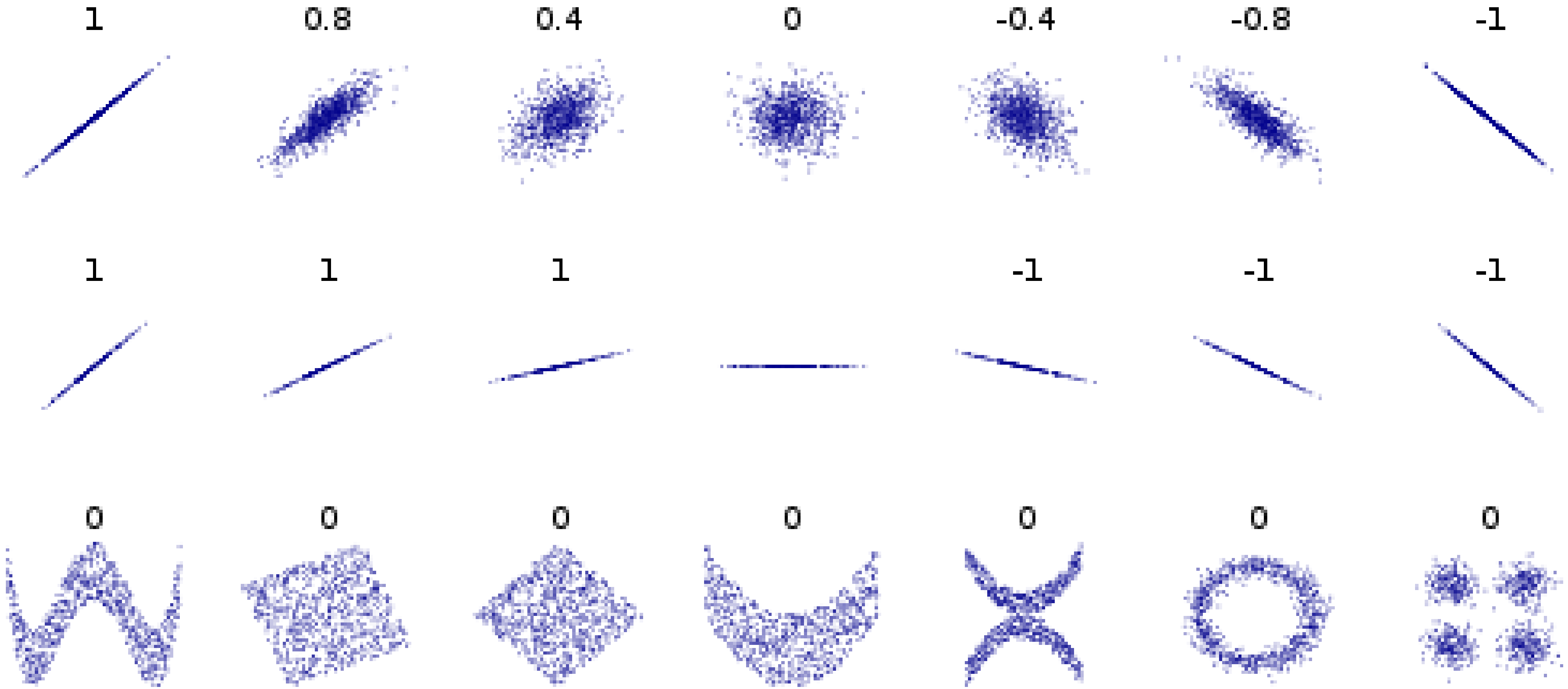
where:

- n is the sample size
- x_i, y_i are the single samples indexed with i
- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (the sample mean); and analogously for $\bar{y}$

Pearson correlation: widely-used measure of the linear **correlation** between variables



Examples of Pearson Correlation



Quantification of Correlation between Neural Spike Trains

- Key part of the analysis of experimental data
- Neural coordination is thought to play a key role in
    - information propagation & processing
    - self-organization of the neural system during development

Designing the Appropriate Temporal Correlation Metric

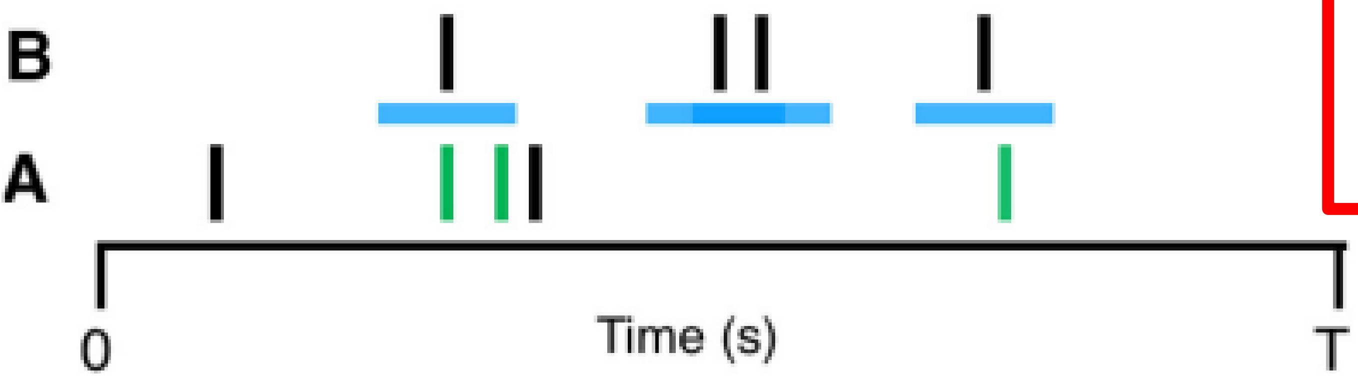
- Symmetry
- Treatment of idle periods
- Robustness to variations in the firing rates

  e.g., doubling the firing rate of two spike trains with a **specific firing structure**, does their correlation remain the same?
- Robust to the recording duration
- Bounded
- Distinction of the correlation vs. no correlation vs. anti-correlation
- Minimal assumptions on the underlying structure/distribution of the events

T_A: the proportion of total recording time which lies within $\pm\Delta t$ of any spike from A. **T_B** calculated similarly.

T_A is given by the fraction of the total recording time (black) which is covered (tiled) by blue bars. Here T_A is 1/3.

P_A: the proportion of spikes from A which lie within $\pm\Delta t$ of any spike from B. **P_B** calculated similarly.



P_A is the number of green spikes in A (3) divided by the total number of spikes in A (5). Here P_A is 3/5.

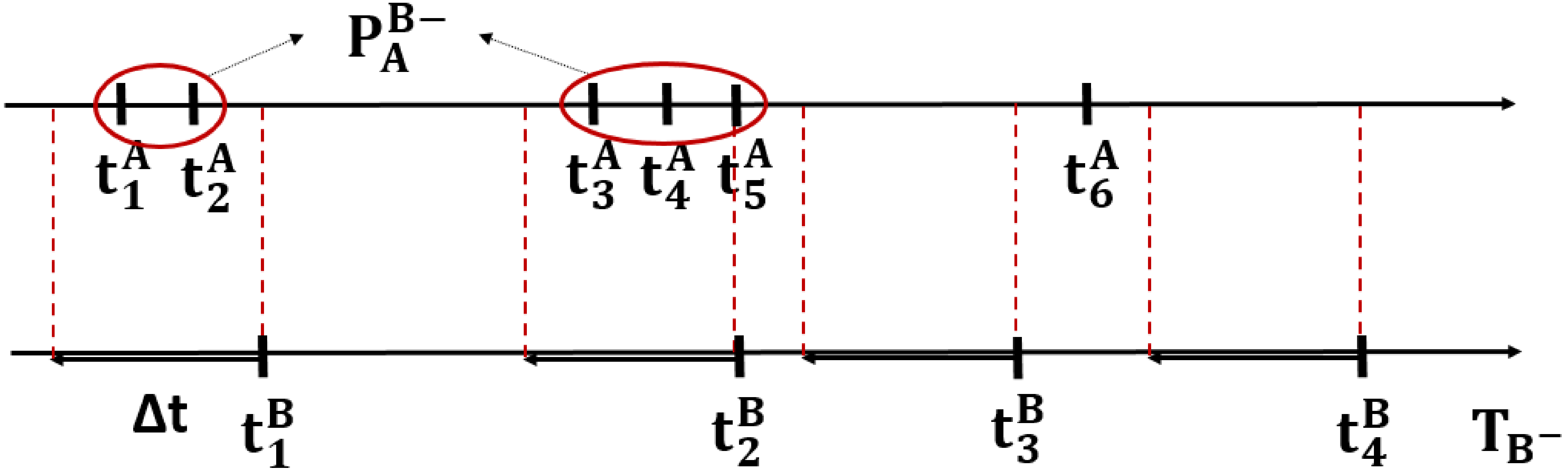
$$STTC = \frac{1}{2}\left(\frac{P_A - T_B}{1 - P_A T_B} + \frac{P_B - T_A}{1 - P_B T_A}\right)$$

Directional STTC
Temporal Correlation Metric

Extended STTC metric to take into consideration the **order** of the correlation of the spike trains of two neurons

Directional STTC$_{AB}$ represents a measure of the chance that firing events of A will **precede** firing events of B

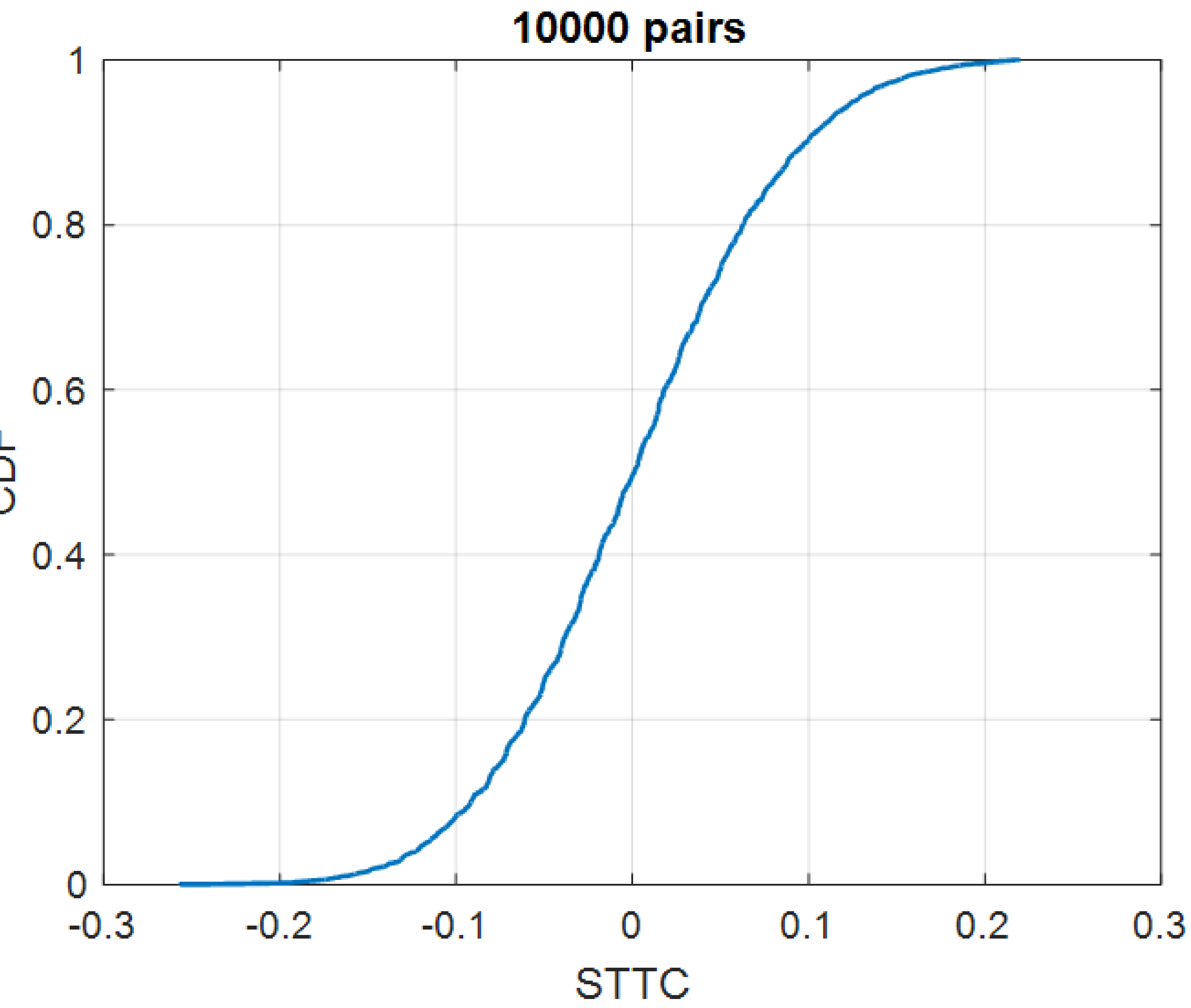
$$STTC_{AB} = \frac{1}{2}\left(\frac{P_A^{B-} - T_{B^-}}{1 - P_A^{B-} T_{B^-}} + \frac{P_B^{A+} - T_{A^+}}{1 - P_B^{A+} T_{A^+}}\right)$$



P_A^{B-}: fraction of firing events of A that occur within an interval Δt prior to firing events of B

T_{B^-}: fraction of total recording time covered by the intervals Δt **prior to each spike of B**

Δt: specific lag (input in directional STTC)



Directional STTC
Synchronous (lag = 0)

Spike trains of 100 time unit
with uniform distr [10, 30] spikes
10,000 pairs

Advantages of Directional STTC

$$STTC_{AB} = \frac{1}{2}\left(\frac{P_A^{B-} - T_{B^-}}{1 - P_A^{B-} T_{B^-}} + \frac{P_B^{A+} - T_{A^+}}{1 - P_B^{A+} T_{A^+}}\right)$$

Relative spike-time shifts (lag parameter)

Order between neurons with respect to their firing events

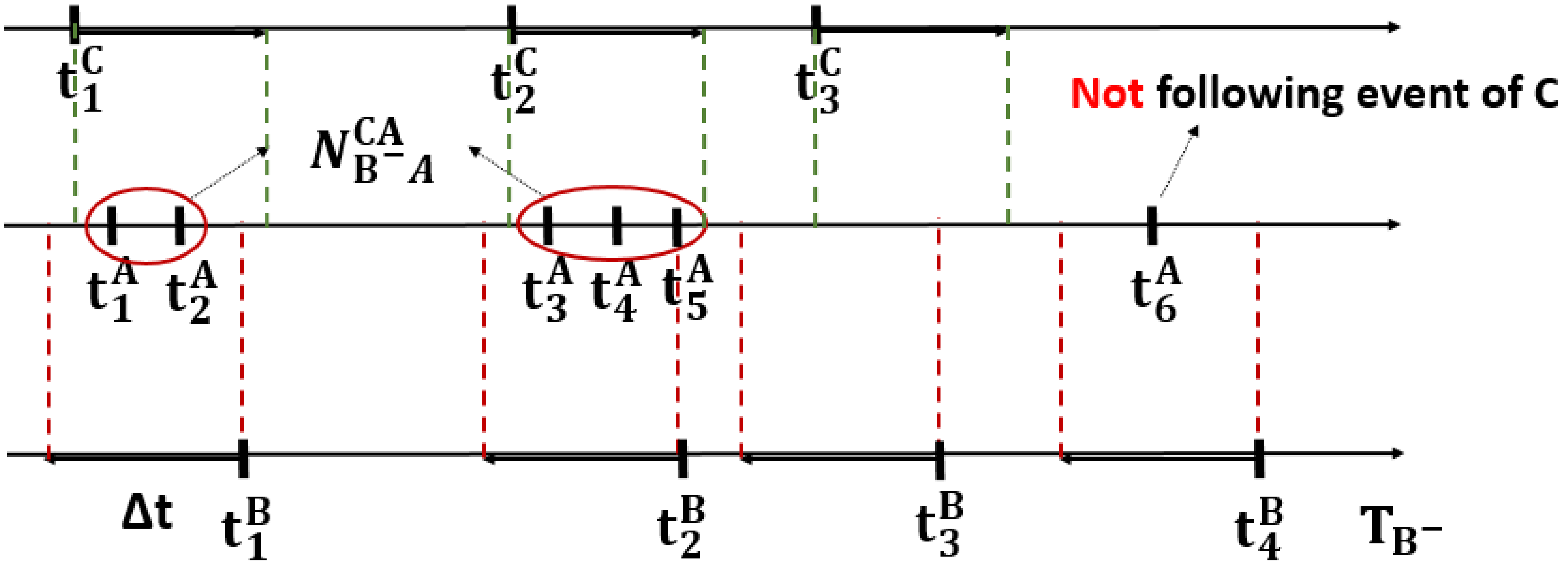
Local fluctuations of neural activity or noise

- accounting the amount of correlation expected by chance

The presence of periods without firing events

- only the firing events contribute

Conditional STTC (A->B |C) represents a measure of the chance that firing events of A will **precede** firing events of B, **given the presence** of firing of C



CA

Conditional STTC (A->B |C) $STTC_{AB}^{C}$

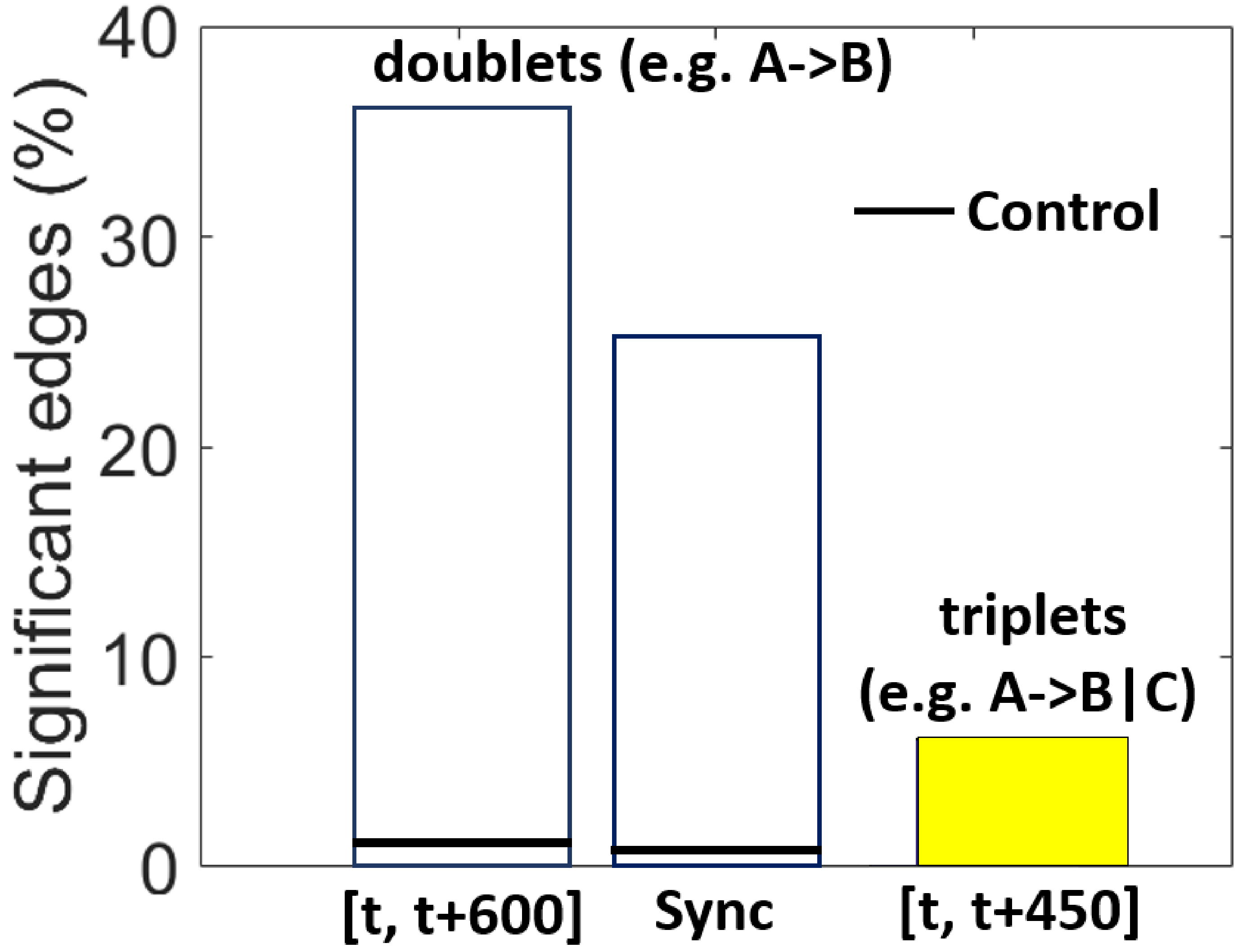
$$STTC_{AB}^{C} = \frac{1}{2}\left(\frac{\frac{N_{B-A}^{CA}}{N_A} - T_{B-}}{1 - \frac{N_{B-A}^{CA}}{N_A}T_{B-}} + \frac{\frac{N_{A+B}^{CA}}{N_B} - T_{A+}}{1 - \frac{N_{A+B}^{CA}}{N_B}T_{A+}}\right)$$

$\mathbf{N_A}$ is the number of firing event in A & $\mathbf{N_B}$ is the number of firing event in B.

$\boldsymbol{T_{A+}}$ is the fraction of the total recording time which is covered by the tiles +Δt after each spike of A, that fall within the tiles Δt after each spike of C.

$\boldsymbol{T_{B-}}$ is the fraction of the total recording time which is covered by the tiles Δt before each spike of B.

Significant Motifs



Significant edge: real STTC value > 3 std. dev. of null distribution

Null distribution: STTC values for the circular shifted neurons (by random delays)

Control (synthetic data)
Each neuron trace is circular shifted by random delay
For each pair of ‘shifted’ neurons, estimate the directional STTC & null distr.
Identify the significant edges

“A→B” indicates that firing events of **A proceed firing events of B** by a specific lag

Null distribution test for directional STTC

For a given pair (A,B)

1. Circular shift the spike train of the neuron A (generated spike train A^i)
2. Estimate the directional STTC(A^i, B)

Repeat the above steps 100 times (i=1, … , 100)

3. Estimate the mean & standard deviation of the obtained STTC values
4. The statistical significant threshold (thr) = mean + 3 std dev

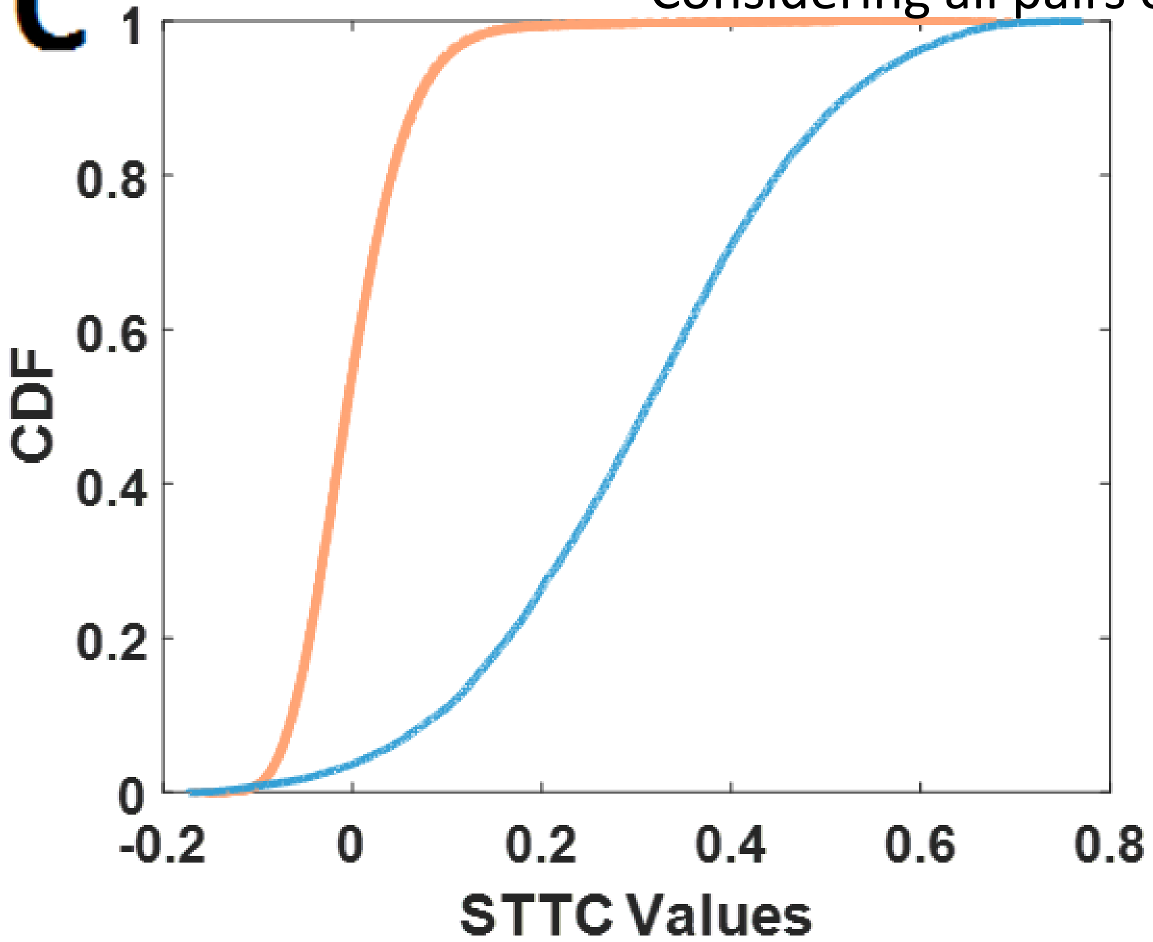
Criterion:

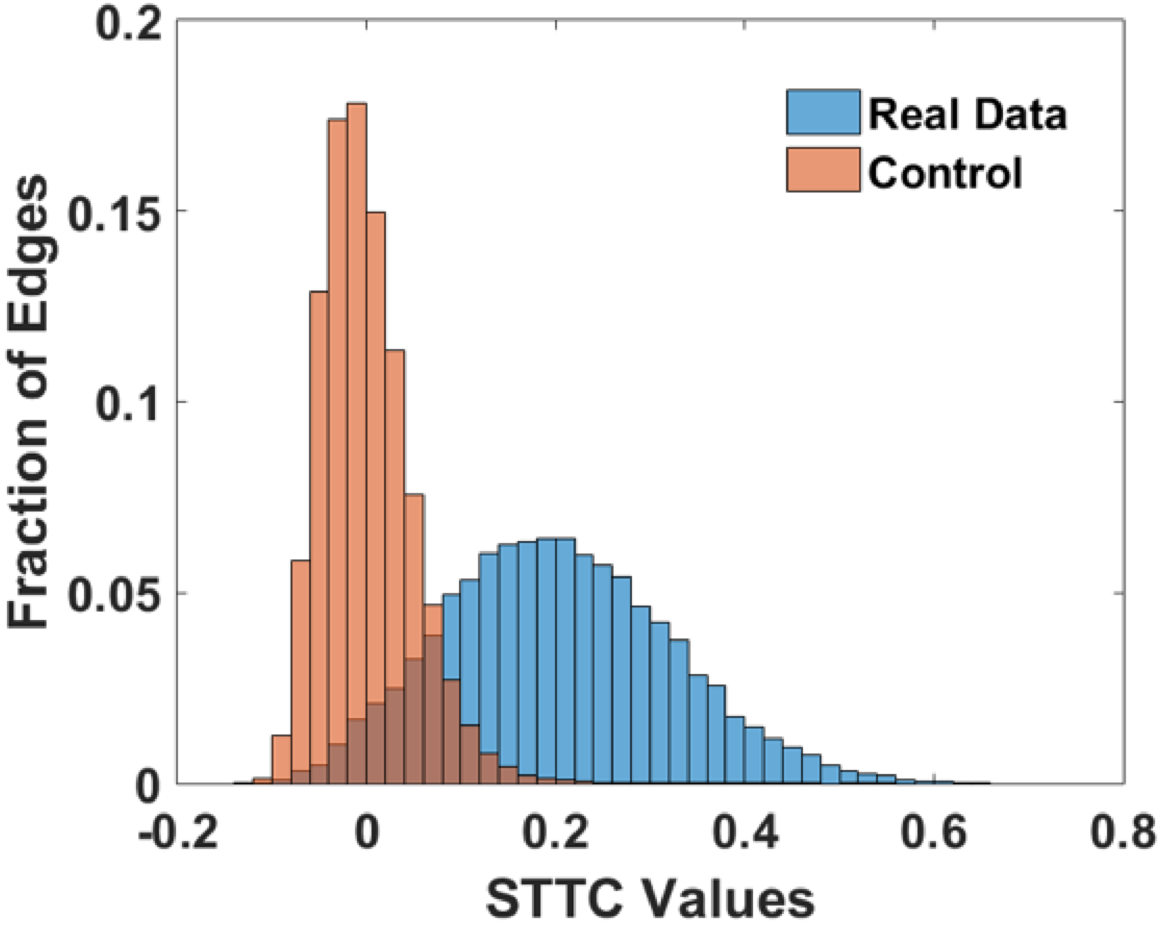
If the directional STTC (A, B) > thr , the directional STTC (A,B) is statistically significant.

The criterion can be strengthen with more repetitions (e.g., **1000)**, a larger number of std dev (e.g., **5**).

C

Considering all pairs of neurons in the datasets





Control group

Each neuron trace is circular shifted by random delay

For each pair of neurons, estimate the directional STTC & null distribution

Identify the significant edges

The **real neuron traces** appear **higher** values of directional STTC & percentage of significant edges

Strengthen the Criterion of Significant Directional STTC (A,B)

Additional requirements

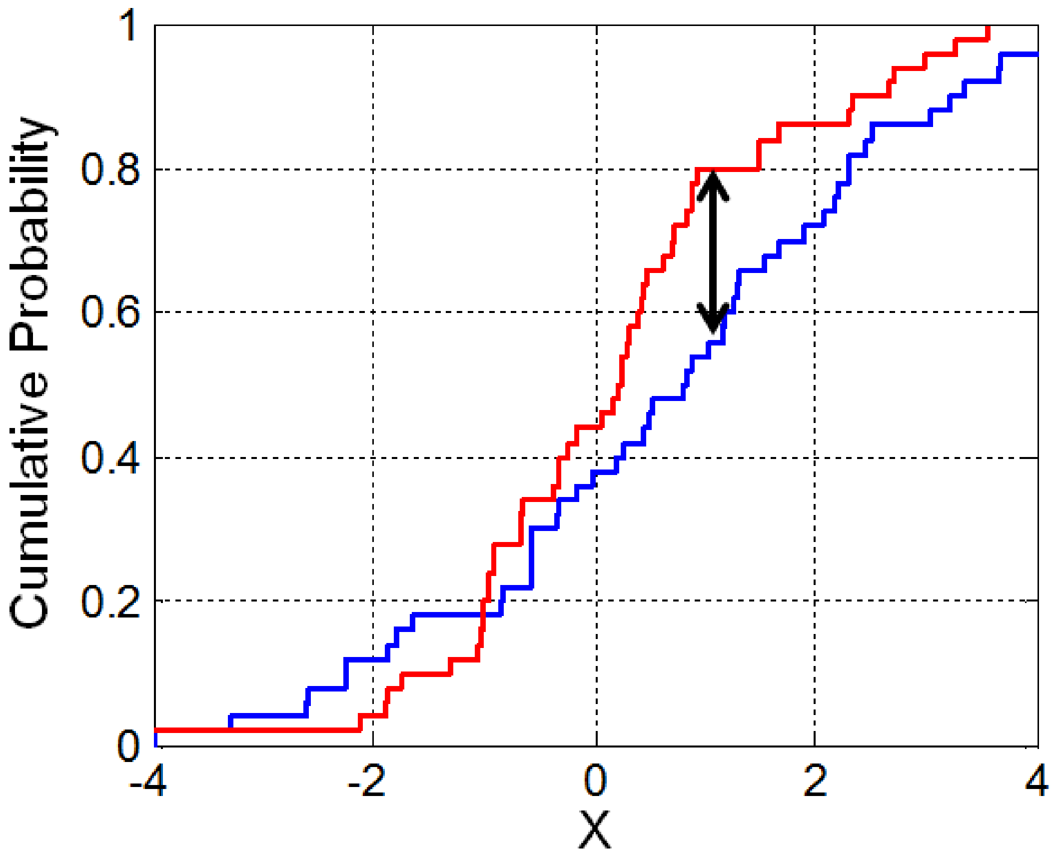
- The total number of spikes of A within a STTC lag of spikes of B is above 3.
- The total number of spikes of B within a STTC lag of spikes of A is above 3.

Kolmogorov-Smirnov (K-S) Test

- **Non-parametric** test of the equality of **continuous 1D** probability distributions
- Quantifies a **distance between two distribution** functions
- Can serve as a **goodness of fit test**

- **Null hypothesis**
  H_0: Two samples drawn from **populations with same distribution**

The maximum absolute difference between the two CDFs



Kolmogorov-Smirnov (K-S) Test

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical distribution fun

The null hypothesis is rejected at level α if

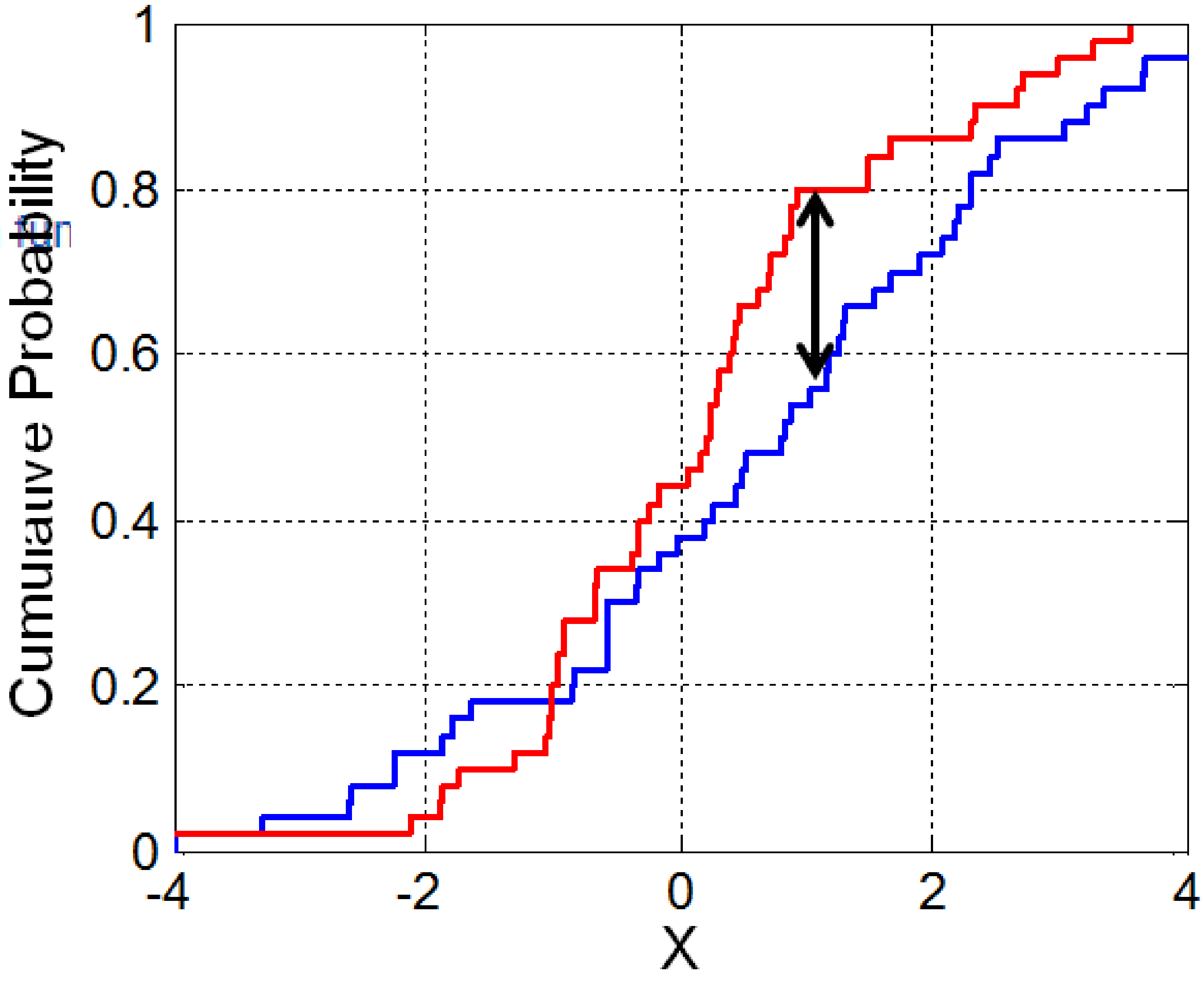
$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{nm}}.$$

n & m: size of the sample datasets

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

and in general by

$$c(\alpha) = \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}.$$



Kolmogorov-Smirnov (K-S) Test

Kolmogorov computed the expected distribution of the distance of the two CDFs when the null hypothesis is true.

Example: Kolmogorov-Smirnov Test

	Decision		**p-value**		**Distance**	
Lag	**True Null**	**Null Null**	**True Null**	**Null Null**	True Null	Null Null
1	1	0	0	0.5427	0.79	0.0076
2	1	0	0	0.2126	0.78	0.0100
3	1	0	0	0.98485	0.75	0.0043
4	1	0	0	0.9937	0.72	0.0040
5	1	0	0	0.9769	0.68	0.00453

Distance of two distributions in sup norm

For **all neuron pairs (**A, B), populate the following distributions with

Population 1: real STTC of the pair (A,B)

Population 2: random circular shift in one of the two spike trains of (A,B)

Population 3: random circular shift in one of the two spike trains of (A,B)

True Null: Population 1 vs. Population 2

Null Null: Population 2 vs. Polulation 3