Learning From Time

Daniela Witten
Departments of Statistics & Biostatistics
University of Washington

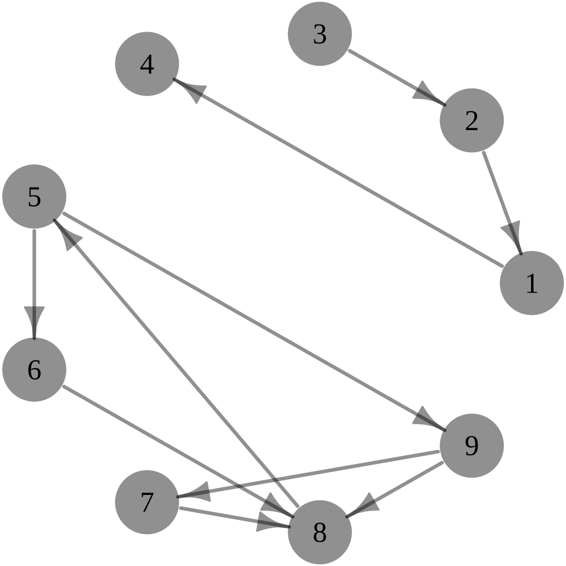
June 9, 2016

Collaborators

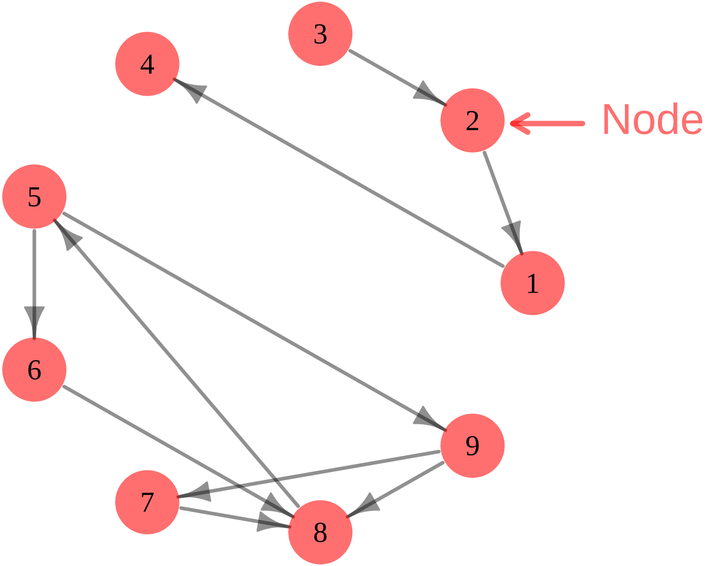
Ali Shojaie

Shizhe Chen

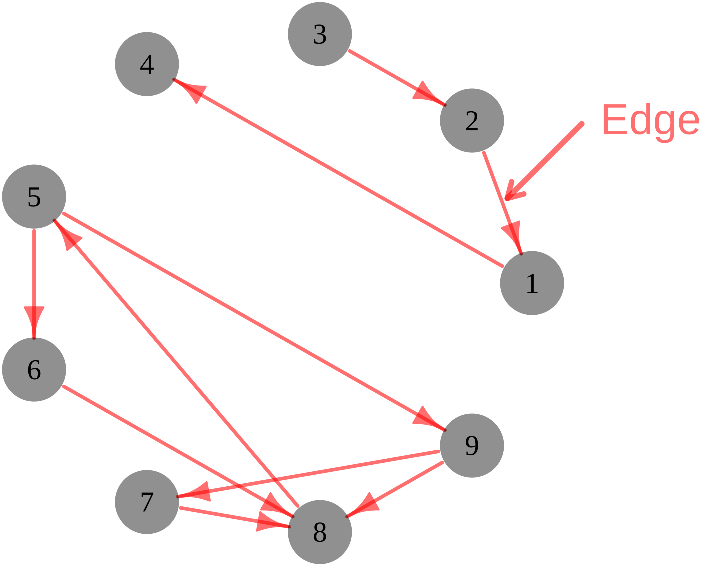
Graphical Model



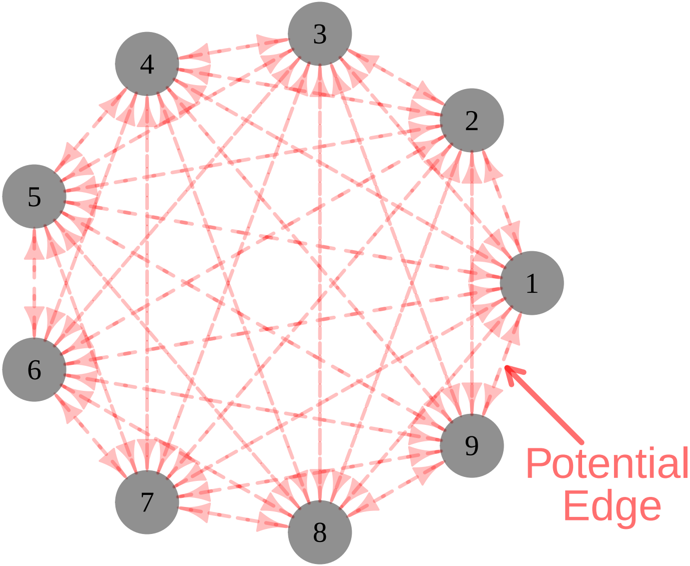
Graphical Model



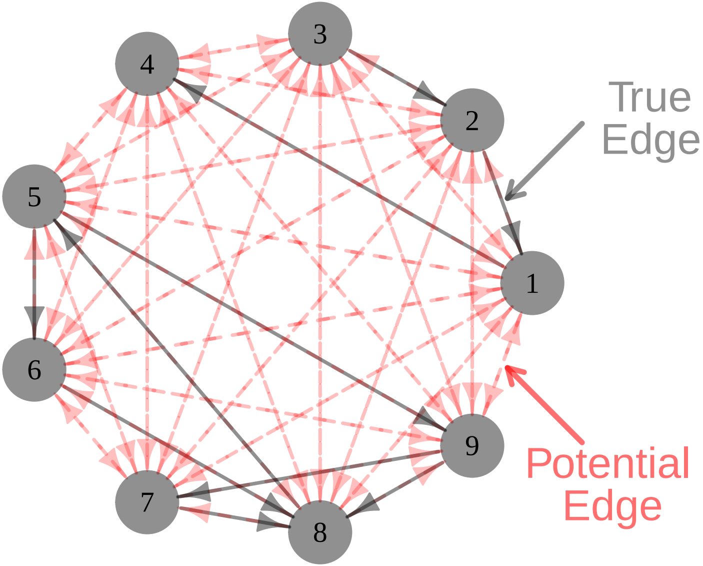
Graphical Model



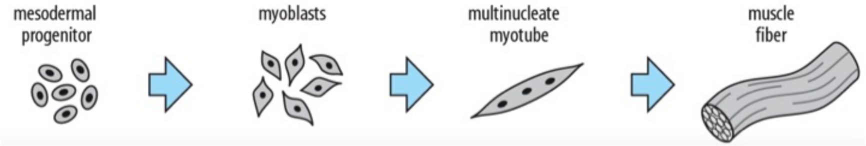
Goal: Learn the Structure of the Graph



Goal: Learn the Structure of the Graph

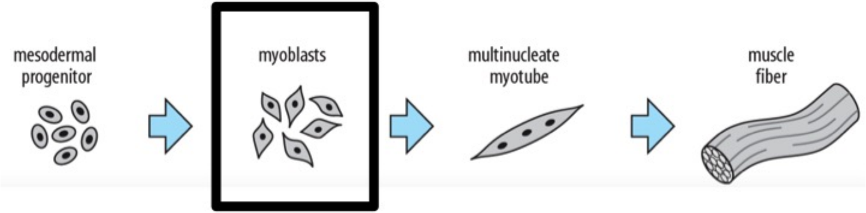


Time is Important

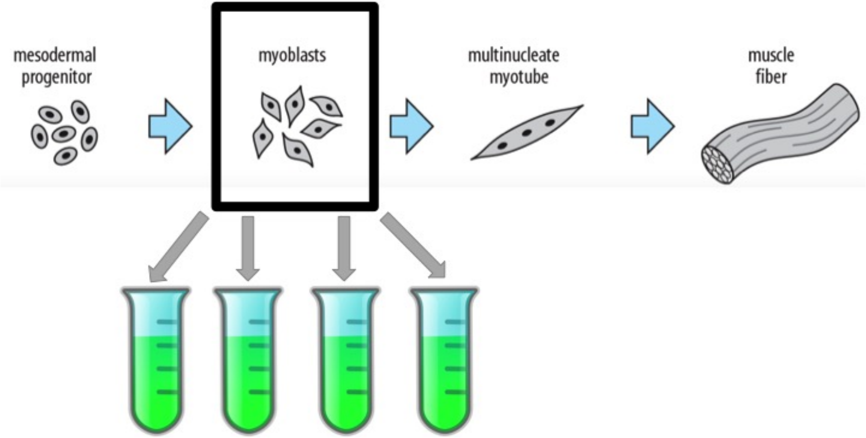


http://www.mun.ca/biology/desmid/brian/BIOL3530/DEVO10/devo10.html

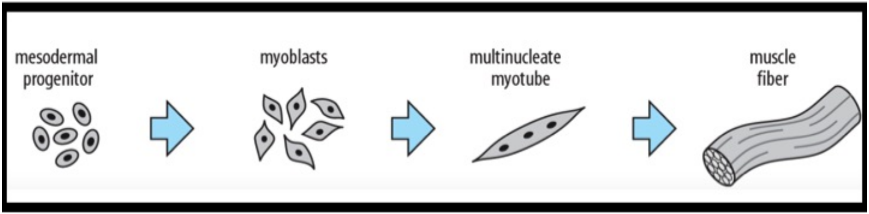
Time is Important



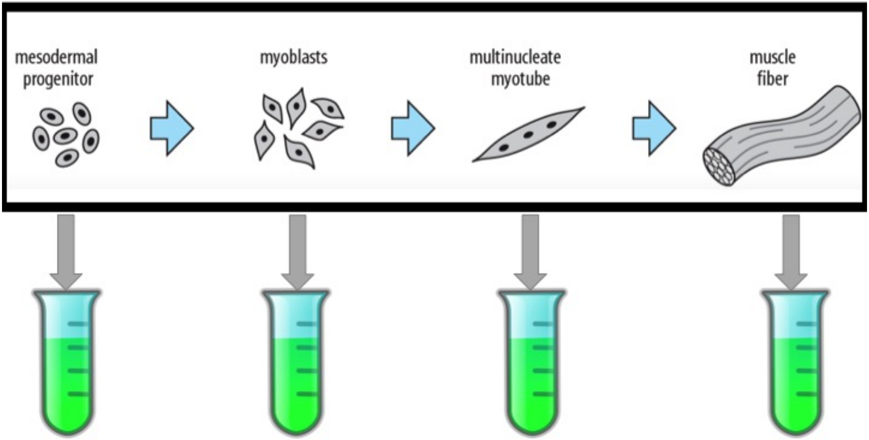
Time is Important



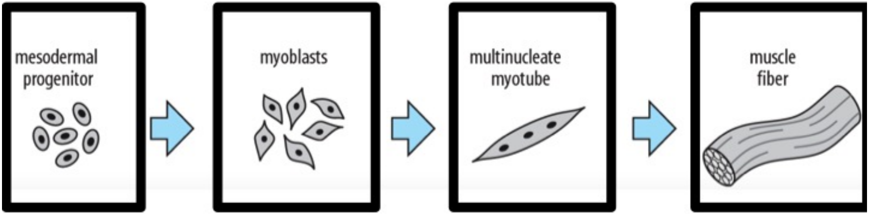
Time is Important



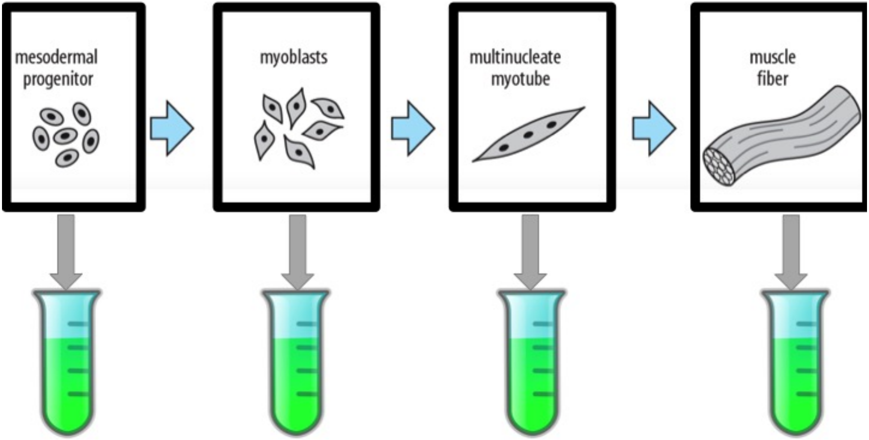
Time is Important



Time is Important



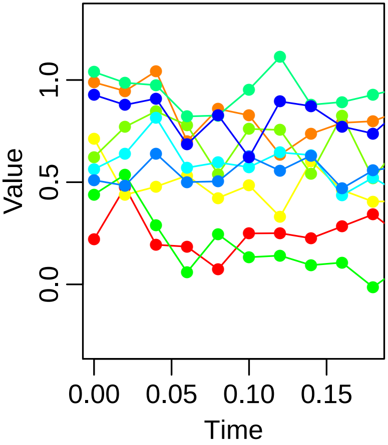
Time is Important



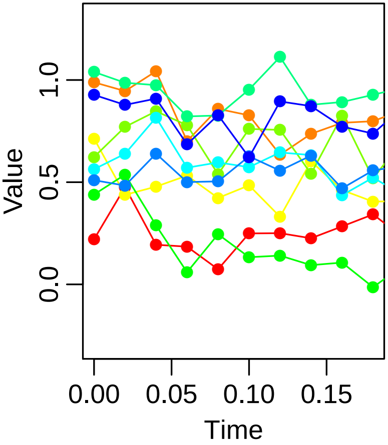
Part I: Learning Gene Regulatory Relationships

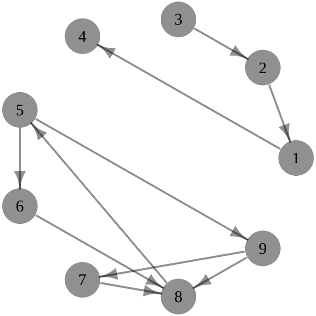
Gene Expression Data

Gene Expression Data

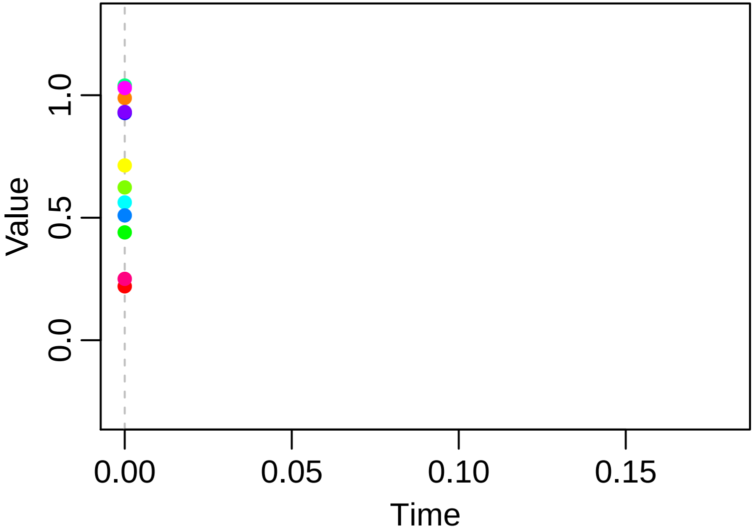


Gene Expression Data

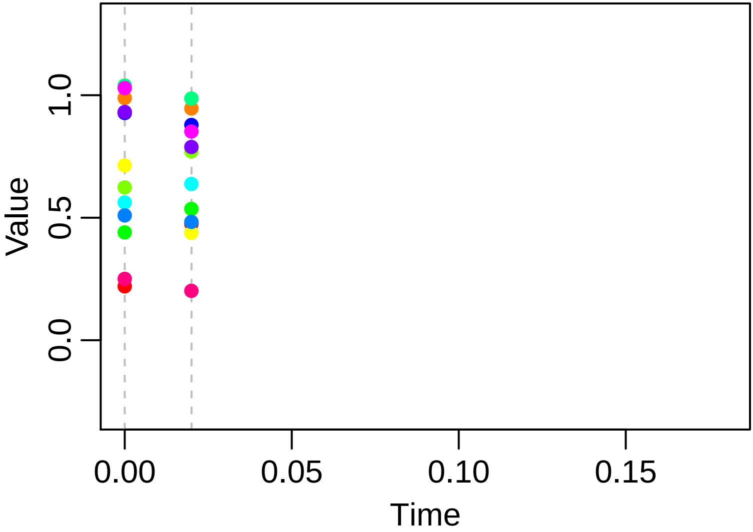




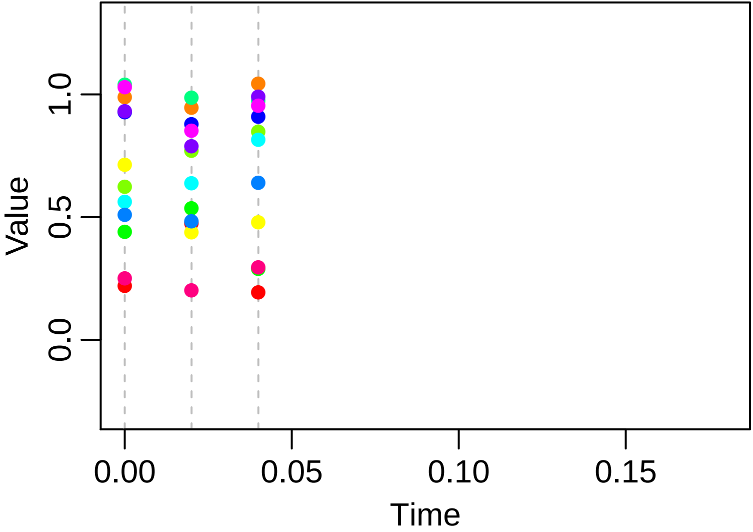
Multivariate Time-Course Data



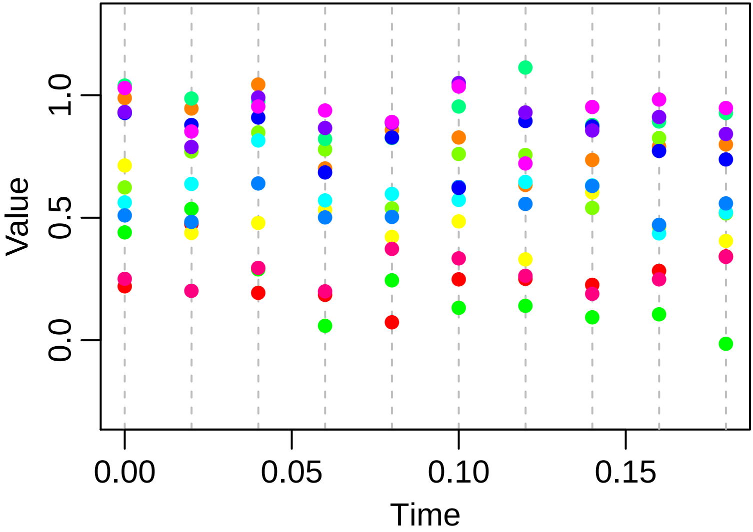
Multivariate Time-Course Data



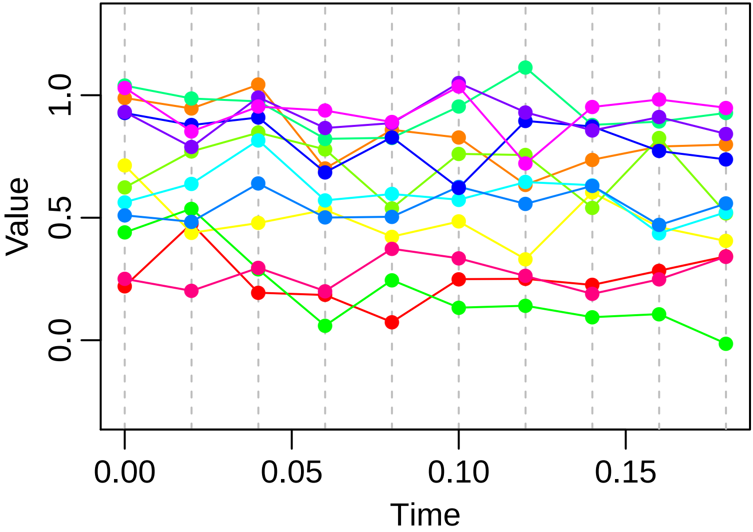
Multivariate Time-Course Data



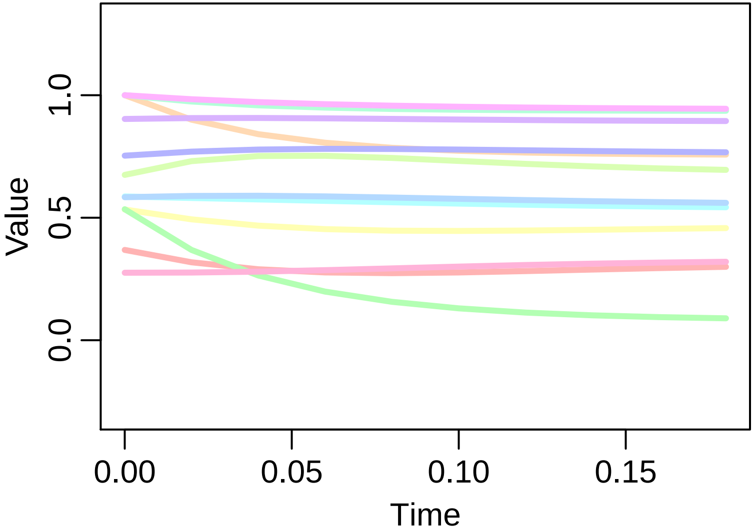
Multivariate Time-Course Data



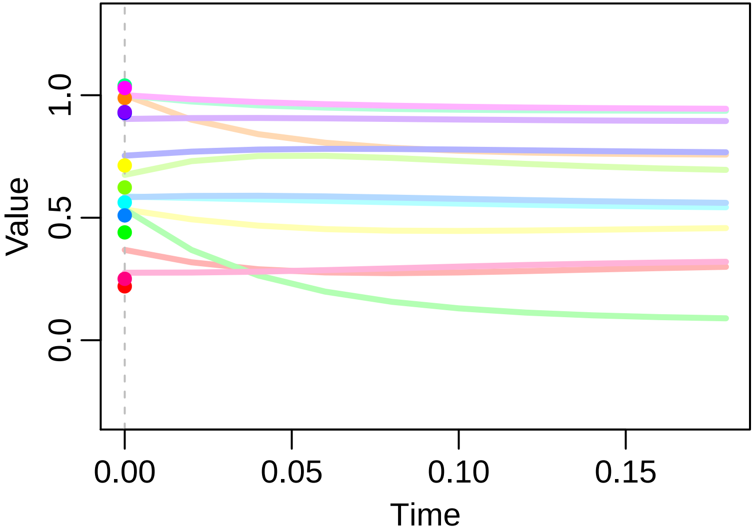
Multivariate Time-Course Data



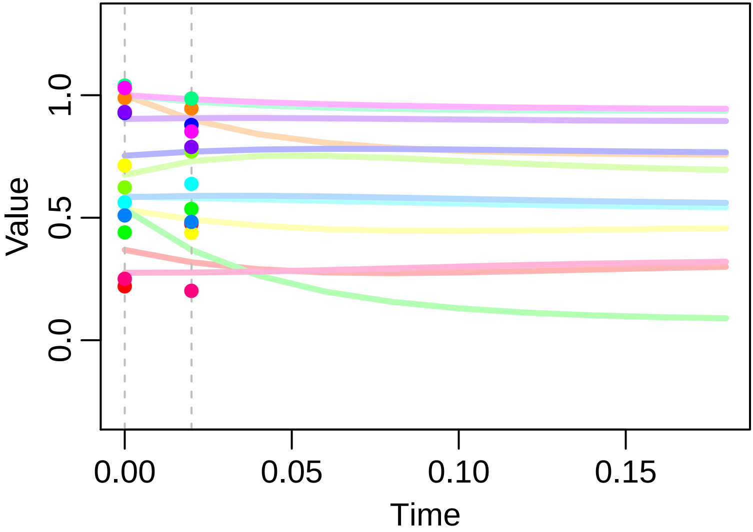
Noiseless Trajectories



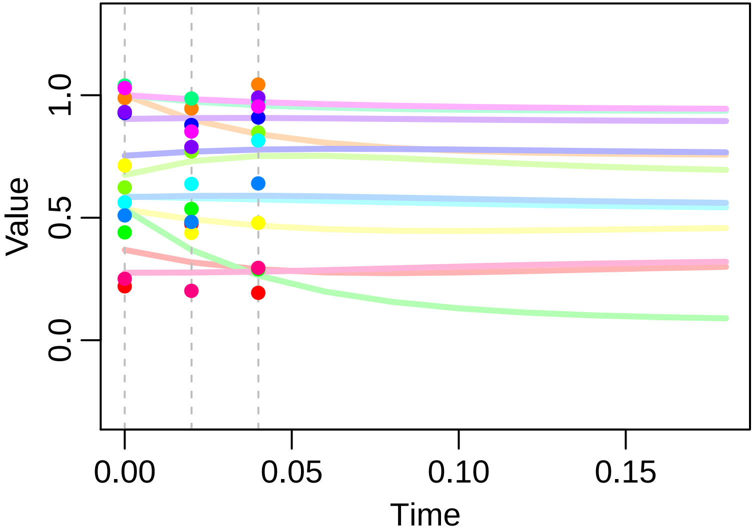
Noiseless Trajectories



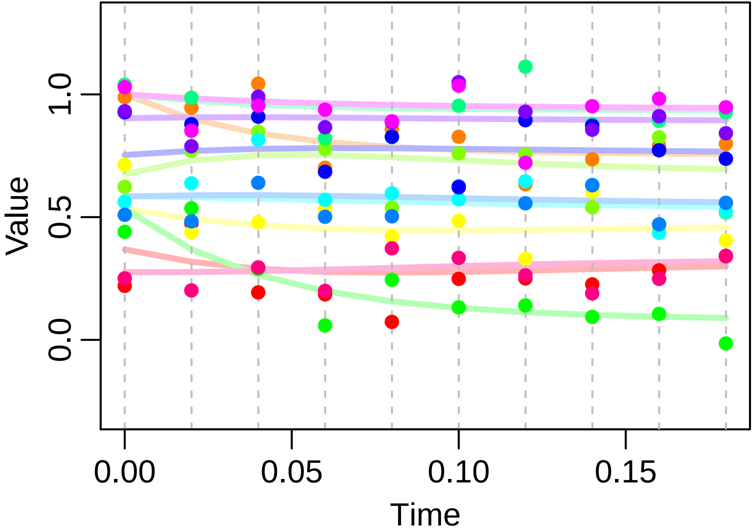
Noiseless Trajectories



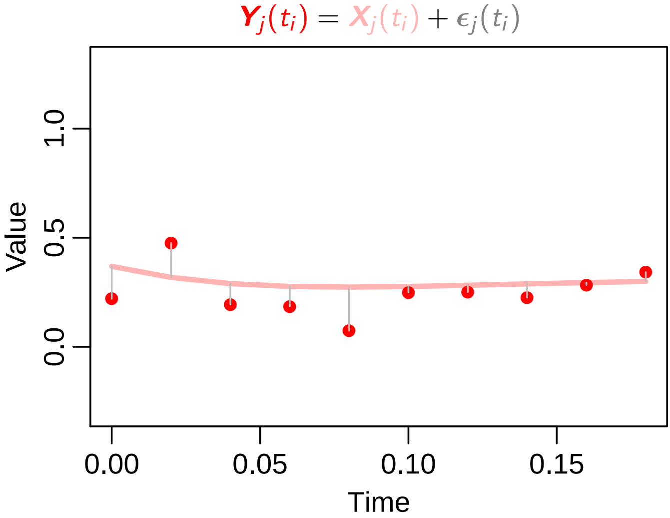
Noiseless Trajectories



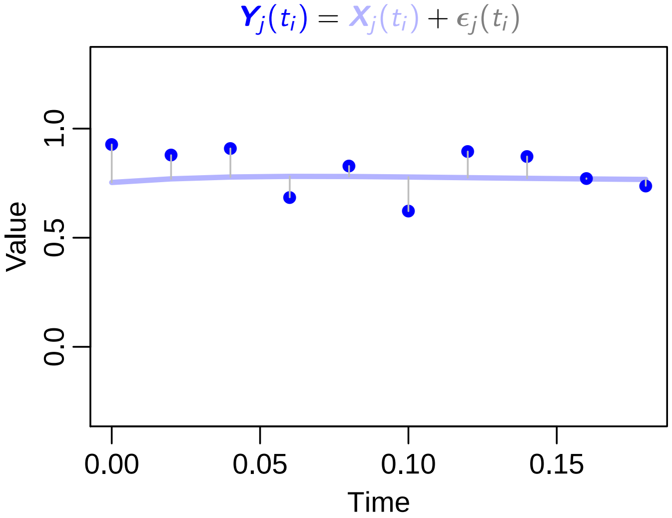
Noiseless Trajectories



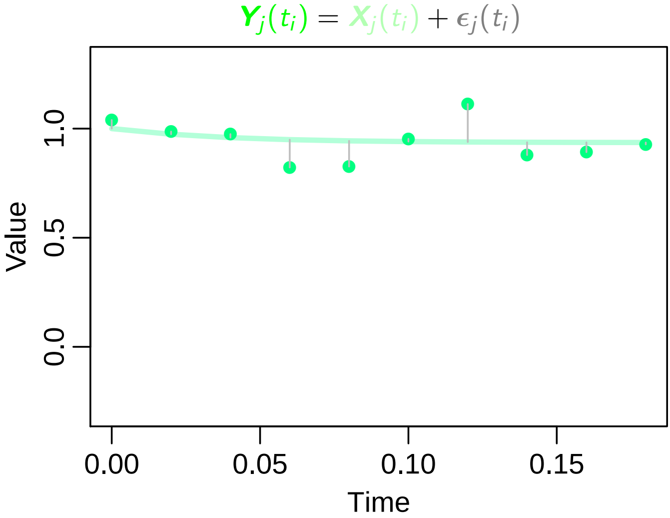
Noiseless Trajectories



Noiseless Trajectories



Noiseless Trajectories



A Model for the Noiseless Trajectories

For $j = 1, \ldots, p$,

$$\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)),$$

where f_{jk} is unknown.

A Model for the Noiseless Trajectories

For $j = 1, \ldots, p$,

$$\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)),$$

where f_{jk} is unknown.

$$\frac{d}{dt} X_1(t) = X_2^2(t) + \exp\left(X_2(t)\right)$$
$$\frac{d}{dt} X_2(t) = 1 + \log\left(X_3(t)\right)$$
$$\frac{d}{dt} X_3(t) = 2$$

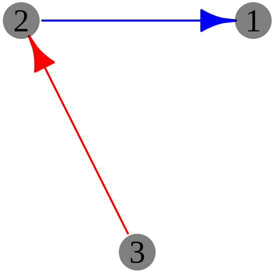
A Model for the Noiseless Trajectories

For $j = 1, \ldots, p$,

$$\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)),$$

where f_{jk} is unknown.

$$\frac{d}{dt} X_1(t) = X_2^2(t) + \exp\left(X_2(t)\right)$$
$$\frac{d}{dt} X_2(t) = 1 + \log\left(X_3(t)\right)$$
$$\frac{d}{dt} X_3(t) = 2$$



Challenges in Fitting the Model, Part I

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t))$$

Ravikumar et al. (2009)

Challenges in Fitting the Model, Part I

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t))$$

Challenge: $f_{jk}(\cdot)$ is unknown.

Ravikumar et al. (2009)

Challenges in Fitting the Model, Part I

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t))$$

Challenge: $f_{jk}(\cdot)$ is unknown.

Solution: Approximate with basis functions, $\psi_1(\cdot), \ldots, \psi_M(\cdot)$:

$$\frac{d}{dt}X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

Ravikumar et al. (2009)

Challenges in Fitting the Model, Part II

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}\left(X_k(t)\right) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

Yuan and Lin (2006); Simon and Tibshirani (2012)

Challenges in Fitting the Model, Part II

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}\left(X_k(t)\right) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

Challenge: $O(Mp^2)$ unknown parameters and N timepoints.

Yuan and Lin (2006); Simon and Tibshirani (2012)

Challenges in Fitting the Model, Part II

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}\left(X_k(t)\right) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

Challenge: $O(Mp^2)$ unknown parameters and N timepoints.

Solution: Group lasso approach to induce sparsity.

Yuan and Lin (2006); Simon and Tibshirani (2012)

Challenges in Fitting the Model, Part III

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}\left(X_k(t)\right) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

Challenges in Fitting the Model, Part III

$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}\left(X_k(t)\right) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

Challenge: $X_k(t)$ is unobserved.

Challenges in Fitting the Model, Part III

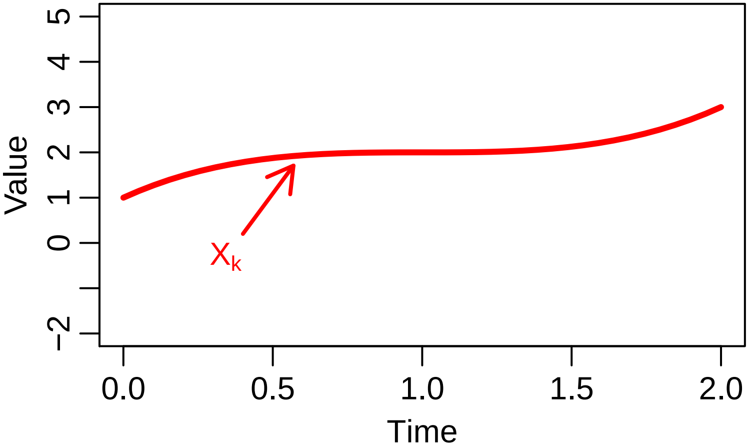
$$\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}\left(X_k(t)\right) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^{\mathrm{T}}\theta_{jk}$$

Challenge: $X_k(t)$ is unobserved.

Solution: Estimate $X_k(t)$ using $Y_k(t_1), \ldots, Y_k(t_N)$.

Existing Methods Estimate the Derivative

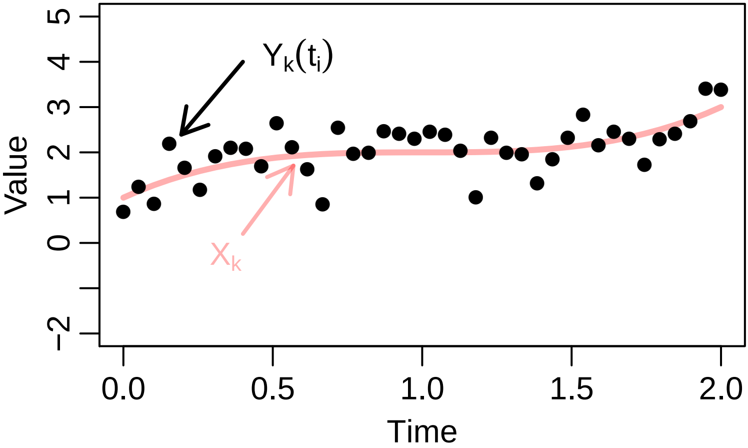
$$\frac{d}{dt}X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t)) \cdot \theta_{jk}$$



Wu et al. (2014) and Henderson and Michailidis (2014)

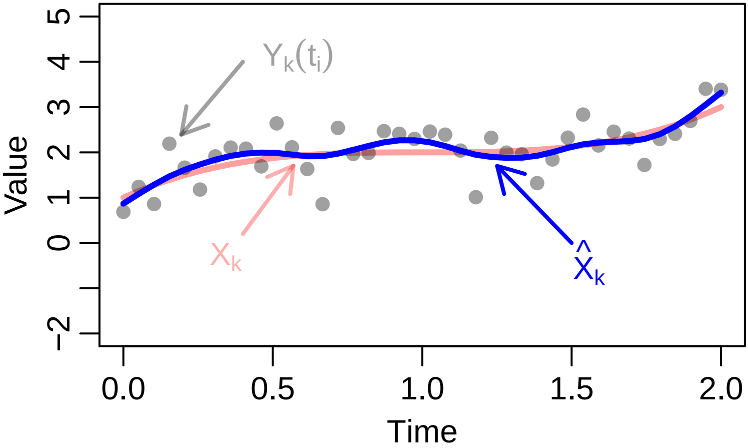
Existing Methods Estimate the Derivative

$$\frac{d}{dt}X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t)) \cdot \theta_{jk}$$



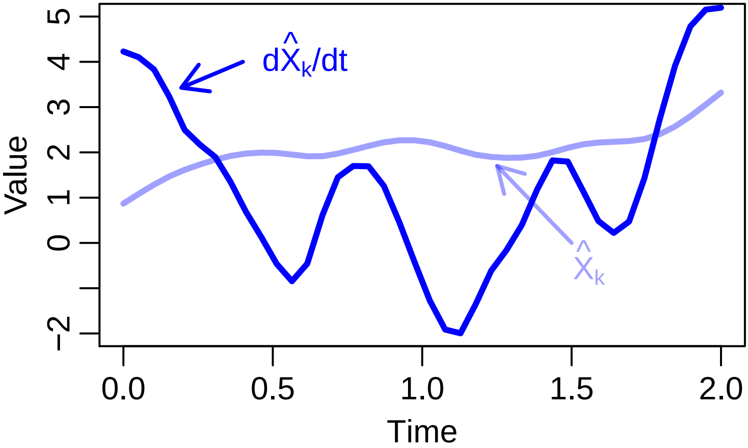
Existing Methods Estimate the Derivative

$$\frac{d}{dt}X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(\hat{X}_k(t)) \cdot \theta_{jk}$$



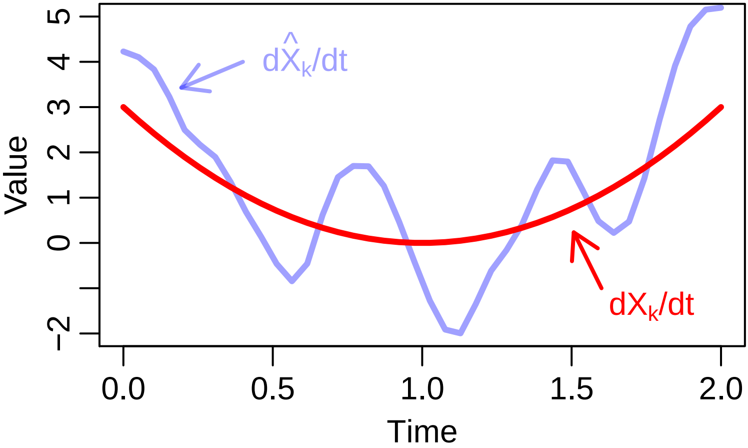
Existing Methods Estimate the Derivative

$$\frac{d}{dt}\hat{X}_j(t) \approx C_j + \sum_{k=1}^{p} \psi(\hat{X}_k(t)) \cdot \theta_{jk}$$



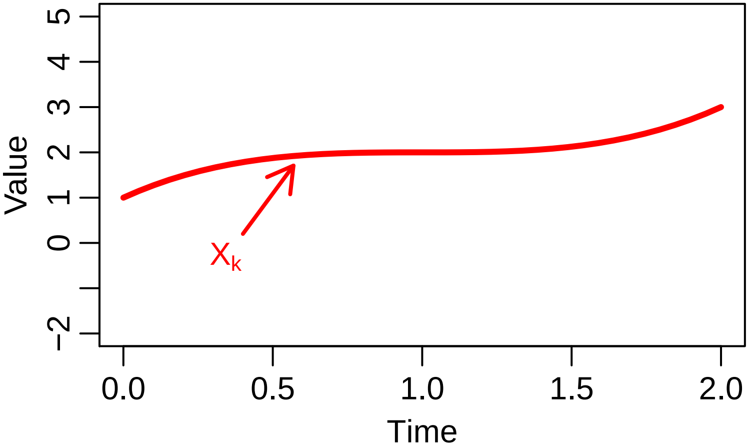
Estimating the Derivative is Hard

$$\frac{d}{dt}\hat{X}_j(t) \text{ and } \frac{d}{dt}X_j(t)$$



Instead, We Can Integrate

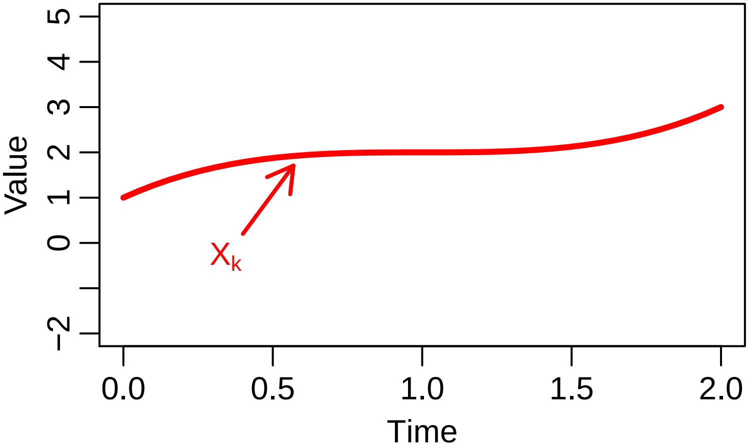
$$\frac{d}{dt}X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t)) \cdot \theta_{jk}$$



The idea of integrating is due to Dattner and Klaassen (2013)

Instead, We Can Integrate

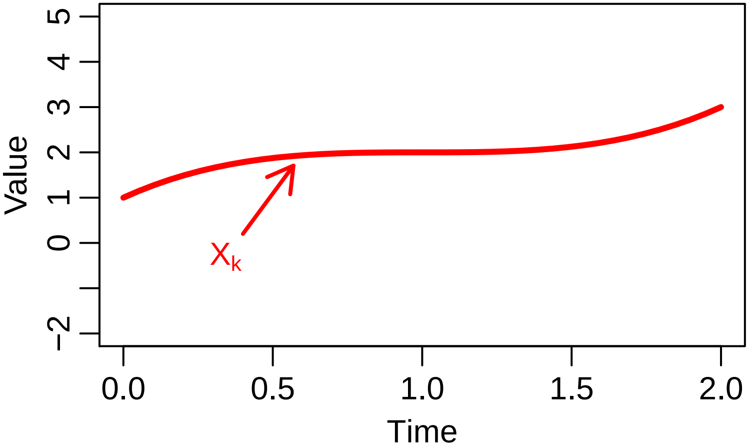
$$\int_0^{t_i} \frac{d}{dt} X_j(s) ds \approx \int_0^{t_i} C_j \, ds + \int_0^{t_i} \sum_{k=1}^{p} \psi(X_k(s)) \cdot \theta_{jk} ds$$



The idea of integrating is due to Dattner and Klaassen (2013)

Instead, We Can Integrate

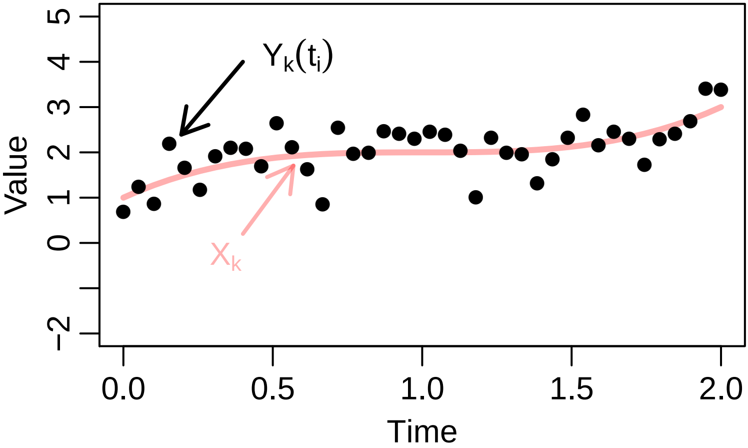
$$X_j(t_i) - X_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[\int_0^{t_i} \psi(X_k(s))ds \right] \cdot \theta_{jk}$$



The idea of integrating is due to Dattner and Klaassen (2013)

Instead, We Can Integrate

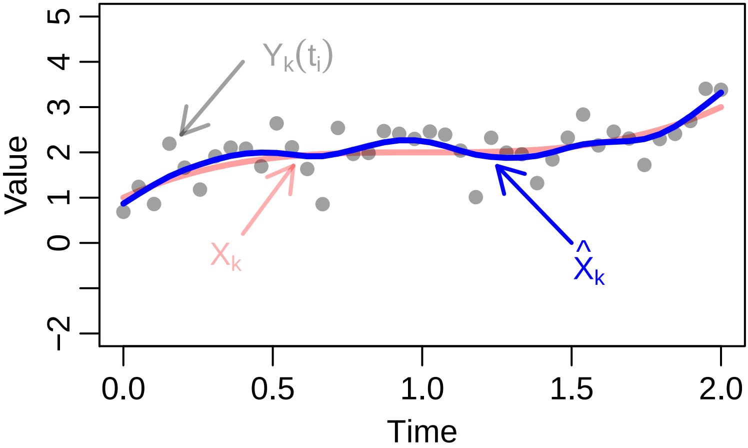
$$Y_j(t_i) - X_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[\int_0^{t_i} \psi(X_k(s)) ds \right] \cdot \theta_{jk}$$



The idea of integrating is due to Dattner and Klaassen (2013)

Instead, We Can Integrate

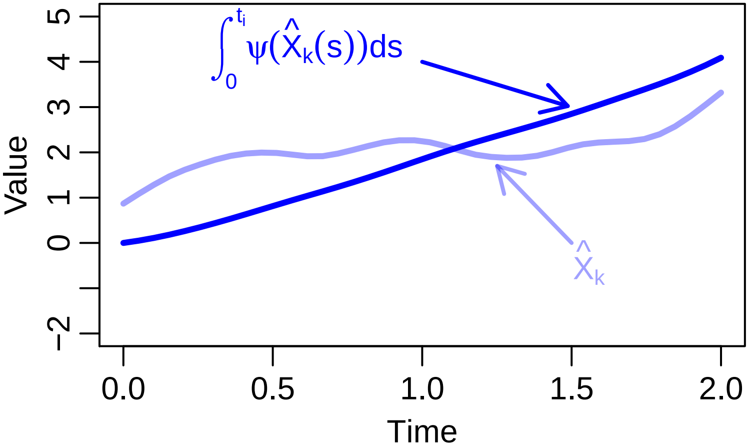
$$Y_j(t_i) - \hat{X}_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[\int_0^{t_i} \psi(X_k(s)) ds \right] \cdot \theta_{jk}$$



The idea of integrating is due to Dattner and Klaassen (2013)

Instead, We Can Integrate

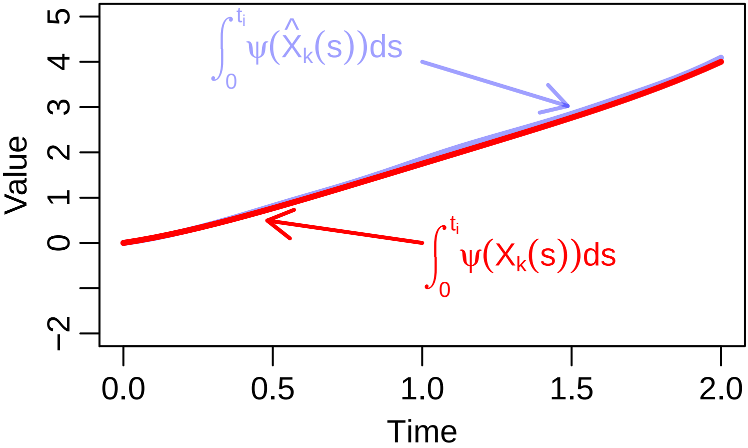
$$Y_j(t_i) - \hat{X}_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[\int_0^{t_i} \psi(\hat{X}_k(s))ds \right] \cdot \theta_{jk}$$



The idea of integrating is due to Dattner and Klaassen (2013)

Estimating the Integral is Easy

$$\int_0^{t_i} \psi(\hat{X}_k(s))\, dt \text{ and } \int_0^{t_i} \psi(X_k(s))\, ds$$



Existing Methods Estimate the Derivative

Wu et al. (2014) and Henderson and Michailidis (2014)

Existing Methods Estimate the Derivative

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\underset{Z(\cdot)\in\chi(h)}{\text{minimize}} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$

Wu et al. (2014) and Henderson and Michailidis (2014)

Existing Methods Estimate the Derivative

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\underset{Z(\cdot)\in\chi(h)}{\text{minimize}} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$

Step 2: For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$\int \left\| \frac{d}{dt}\hat{X}_j(t) - C_j - \sum_{k=1}^{p} \psi(\hat{X}_k(t))^{\mathrm{T}} \theta_{jk} \right\|_2^2 dt$$
$$+ \lambda \sum_{k=1}^{p} \underbrace{\sqrt{\int \left(\psi(\hat{X}_k(t))^{\mathrm{T}} \theta_{jk} \right)^2 dt}}_{\text{standardized group lasso}}.$$

Wu et al. (2014) and Henderson and Michailidis (2014)

Existing Methods Estimate the Derivative

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\underset{Z(\cdot)\in\chi(h)}{\text{minimize}} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$

Step 2: For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$\int \left\| \frac{d}{dt}\hat{X}_j(t) - C_j - \sum_{k=1}^{p} \psi(\hat{X}_k(t))^{\mathrm{T}}\theta_{jk} \right\|_2^2 dt$$
$$+ \lambda \sum_{k=1}^{p} \underbrace{\sqrt{\int \left(\psi(\hat{X}_k(t))^{\mathrm{T}}\theta_{jk} \right)^2 dt}}_{\text{standardized group lasso}}.$$

Step 3: The graph estimate is $\hat{\mathcal{E}} = \left\{ (j,k) : \hat{\theta}_{jk} \neq 0 \right\}$.

Wu et al. (2014) and Henderson and Michailidis (2014)

Our Proposal: **G**raph **R**econstruction w/ **A**dditive **D**ifferential **E**quations

Our Proposal:
Graph **R**econstruction w/ **A**dditive **D**ifferential **E**quations

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\underset{Z(\cdot)\in\chi(h)}{\text{minimize}} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$

Our Proposal:
Graph **R**econstruction w/ **A**dditive **D**ifferential **E**quations

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\underset{Z(\cdot)\in\chi(h)}{\text{minimize}} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$

Step 2: For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$\sum_{i=1}^{n} \left[Y_j(t_i) - \hat{X}_j(0) - t_i C_j - \sum_{k=1}^{p} \hat{\Psi}_{ik}^{\mathrm{T}} \theta_{jk} \right]^2 + \lambda \sum_{k=1}^{p} \underbrace{\sqrt{\sum_{i=1}^{n} \left(\hat{\Psi}_{ik}^{\mathrm{T}} \theta_{jk} \right)^2}}_{\text{standardized group lasso}},$$

where $\hat{\Psi}_{ik} = \int_0^{t_i} \psi(\hat{X}_k(s))\, ds, i = 1, \ldots, n.$

Our Proposal: Graph Reconstruction w/ Additive Differential Equations

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\underset{Z(\cdot)\in\chi(h)}{\text{minimize}} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$

Step 2: For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$\sum_{i=1}^{n} \left[Y_j(t_i) - \hat{X}_j(0) - t_i C_j - \sum_{k=1}^{p} \hat{\Psi}_{ik}^{\mathrm{T}} \theta_{jk} \right]^2 + \lambda \sum_{k=1}^{p} \underbrace{\sqrt{\sum_{i=1}^{n} \left(\hat{\Psi}_{ik}^{\mathrm{T}} \theta_{jk} \right)^2}}_{\text{standardized group lasso}},$$

where $\hat{\Psi}_{ik} = \int_0^{t_i} \psi(\hat{X}_k(s))\, ds, i = 1, \ldots, n.$

Step 3: The graph estimate is $\hat{\mathcal{E}} = \left\{ (j, k) : \hat{\theta}_{jk} \neq 0 \right\}$.

Theory – Overview Of Our Results

- We bound

$$\int_t \left\{ \hat{X}_j(t) - X_j(t) \right\}^2 dt,$$

which allows us to bound $\|\hat{\Psi} - \Psi\|$ in high dimensions.

Extending Loh and Wainwright (2012)

Theory – Overview Of Our Results

- We bound
$$\int_t \left\{\hat{X}_j(t) - X_j(t)\right\}^2 dt,$$
which allows us to bound $\|\hat{\Psi} - \Psi\|$ in high dimensions.
- We establish variable selection consistency of (standardized) group lasso regression with errors-in-variables.

Extending Loh and Wainwright (2012)

Theory – Overview Of Our Results

- We bound

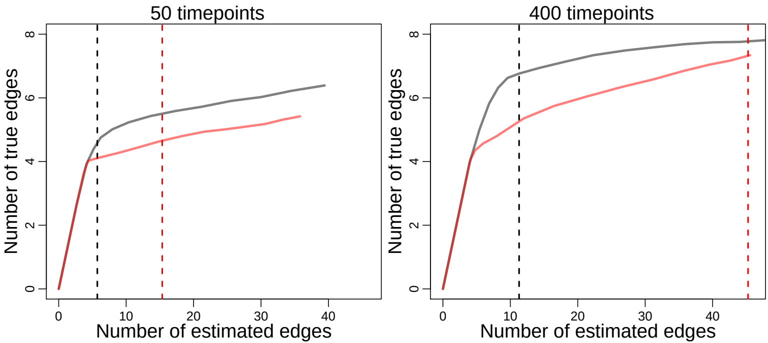
$$\int_t \left\{ \hat{X}_j(t) - X_j(t) \right\}^2 dt,$$

  which allows us to bound $\|\hat{\Psi} - \Psi\|$ in high dimensions.
- We establish variable selection consistency of (standardized) group lasso regression with errors-in-variables.
- We show that with high probability, GRADE correctly identifies the parents of each node.

Extending Loh and Wainwright (2012)

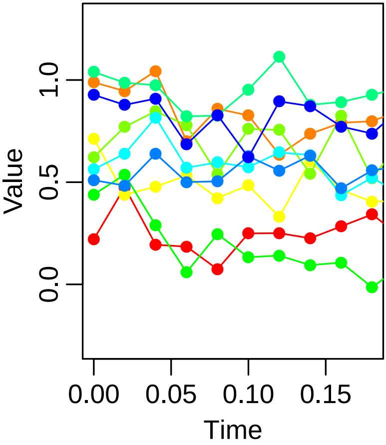
Simulation Results

- NeRDS: Network Reconstruction via Dynamic Systems
- GRADE

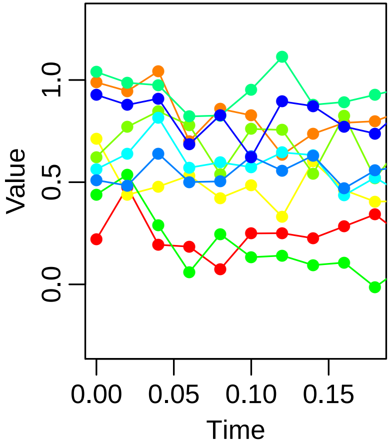


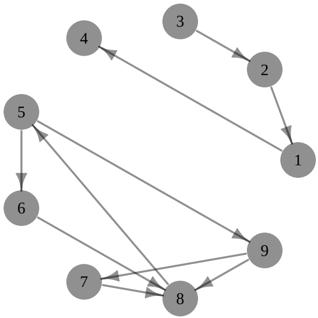
NeRDS is the proposal of Henderson and Michailidis (2014)

The End Result



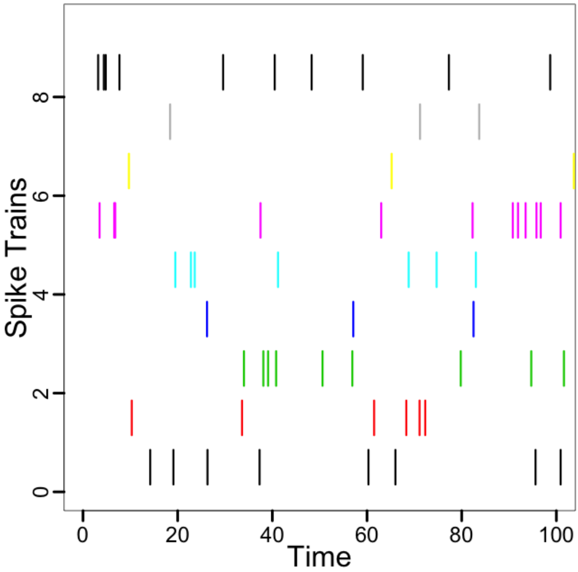
The End Result





Part II: Learning Functional Connectivity Among Neurons

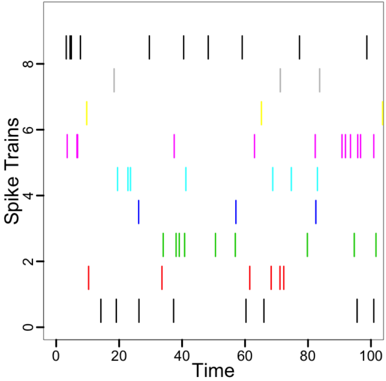
Neuronal Spike Train Data



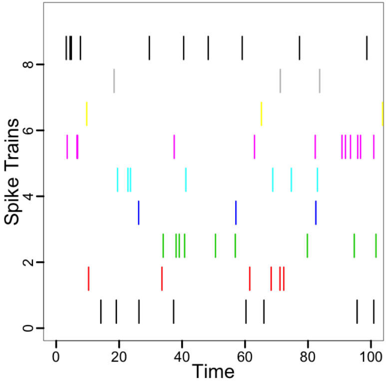
See e.g. Pillow et al. (2008)

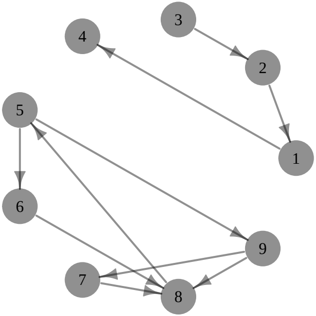
Neuronal Spike Train Data

Neuronal Spike Train Data

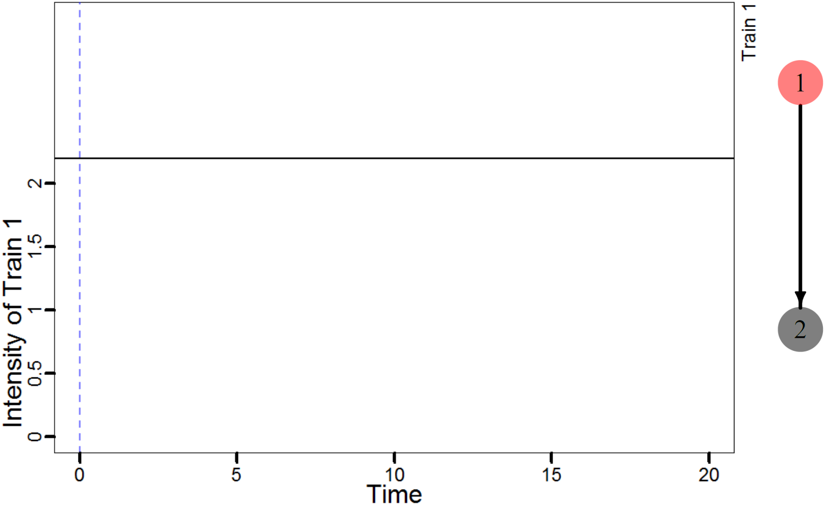


Neuronal Spike Train Data



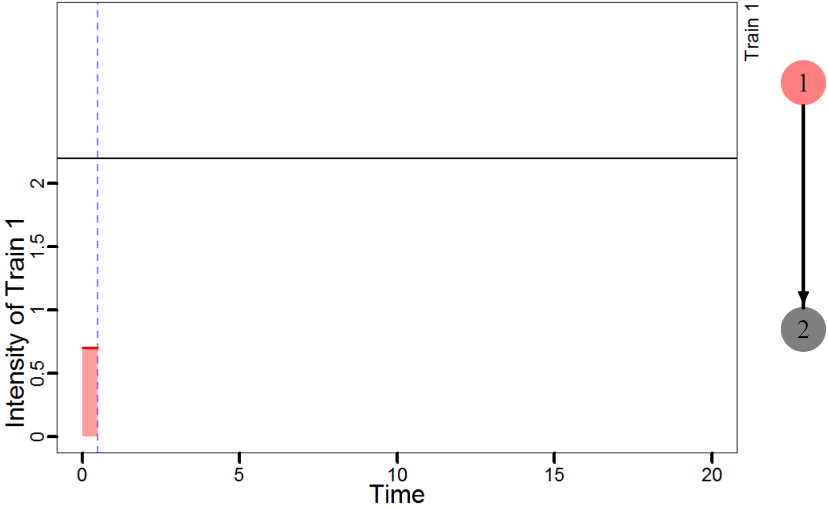


The Hawkes Process



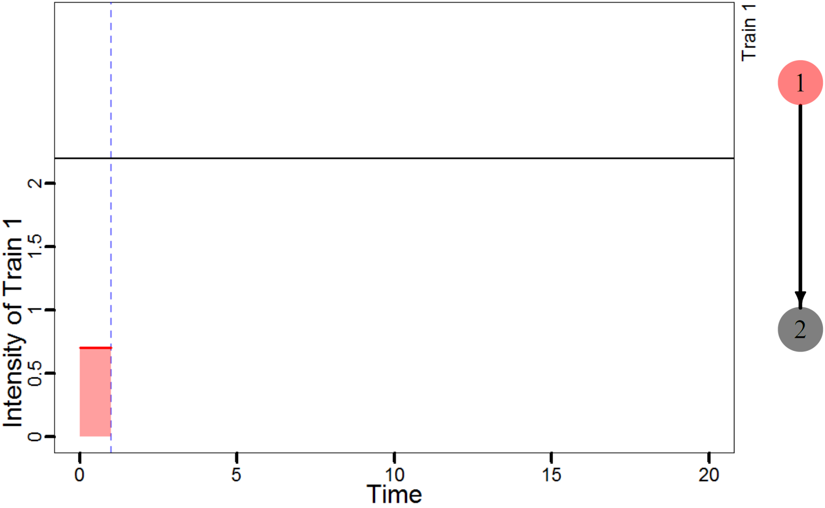
Hawkes (1971)

The Hawkes Process



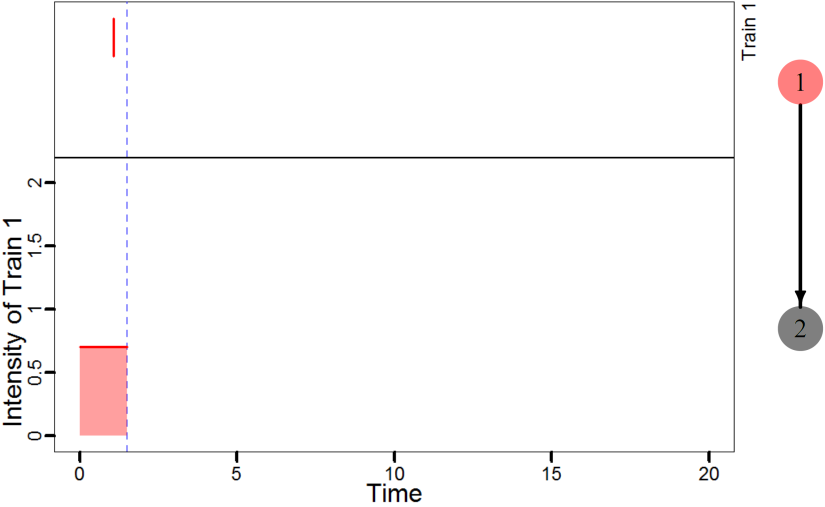
Hawkes (1971)

The Hawkes Process



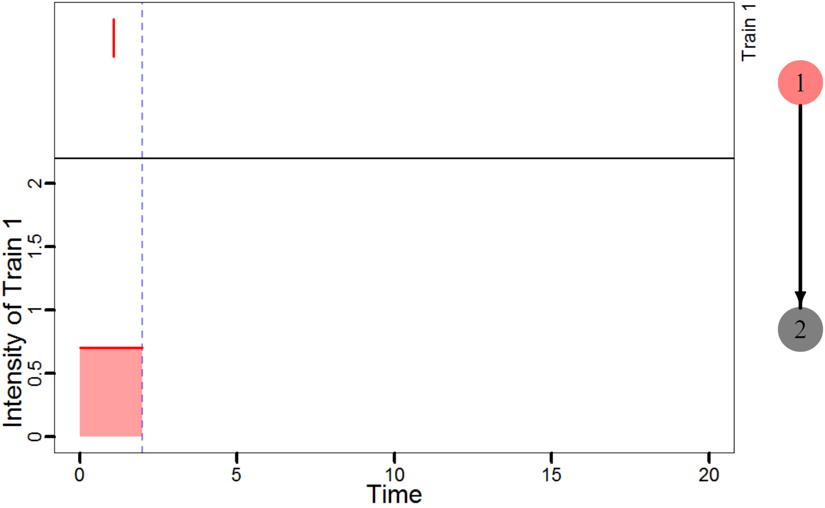
Hawkes (1971)

The Hawkes Process



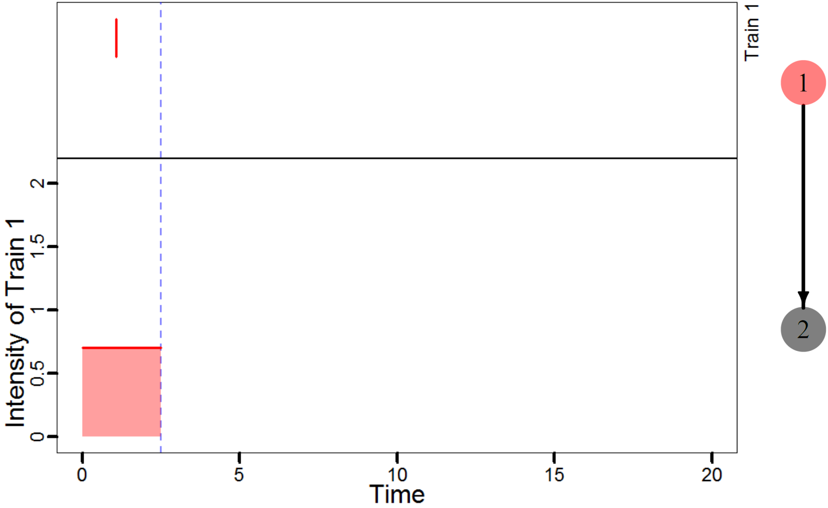
Hawkes (1971)

The Hawkes Process



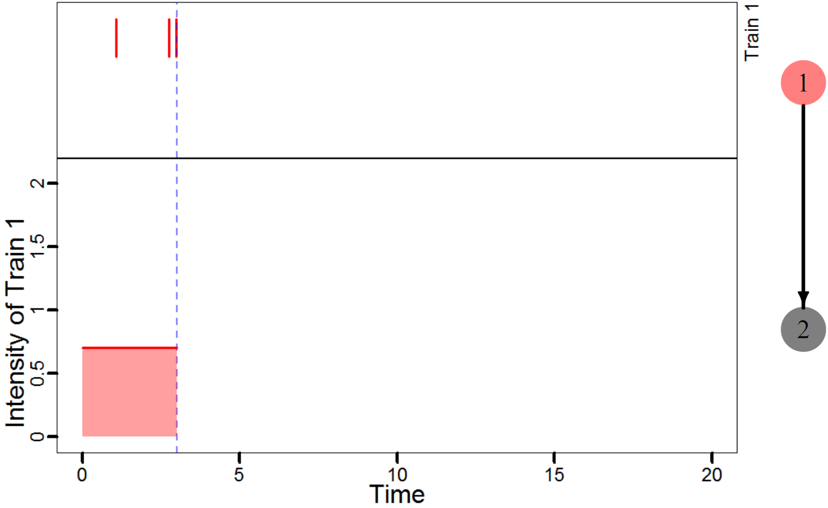
Hawkes (1971)

The Hawkes Process



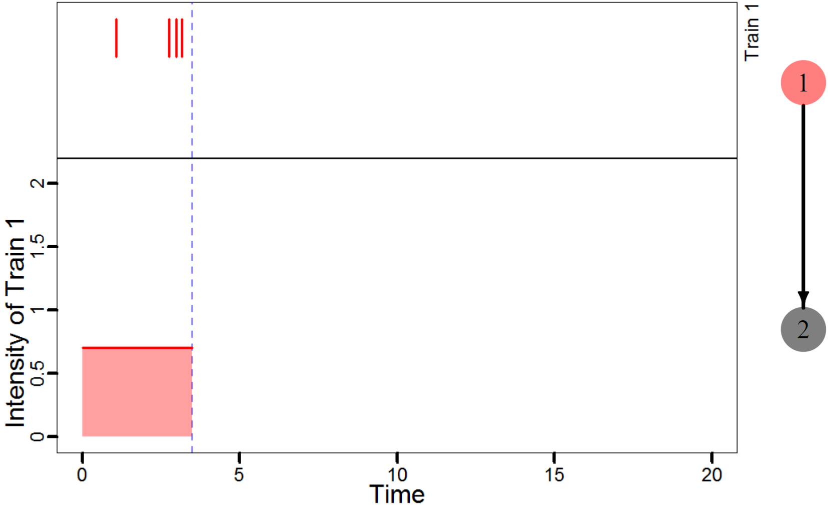
Hawkes (1971)

The Hawkes Process



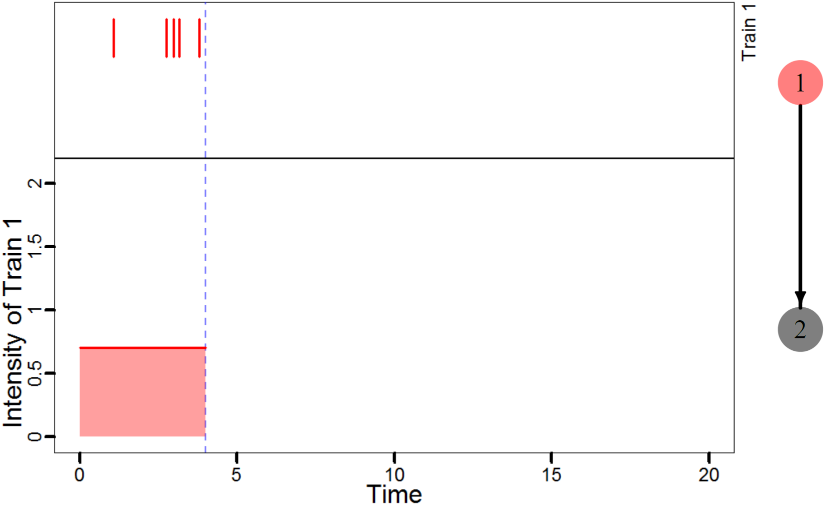
Hawkes (1971)

The Hawkes Process



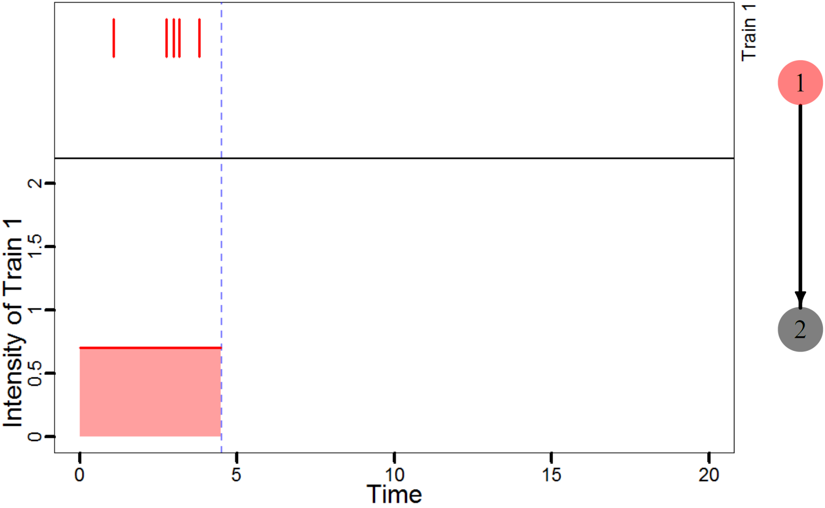
Hawkes (1971)

The Hawkes Process



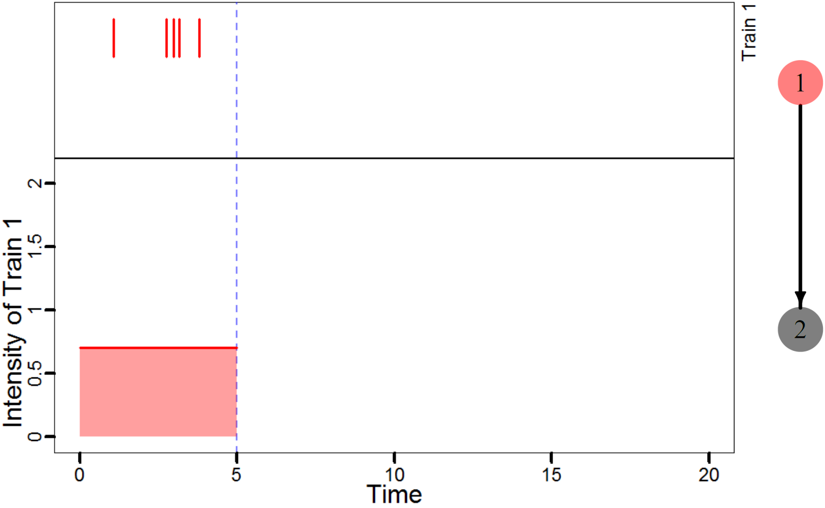
Hawkes (1971)

The Hawkes Process



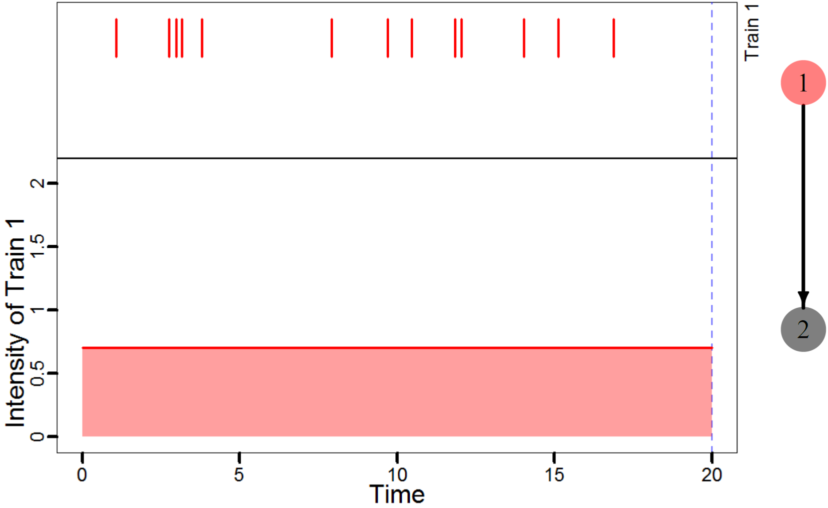
Hawkes (1971)

The Hawkes Process



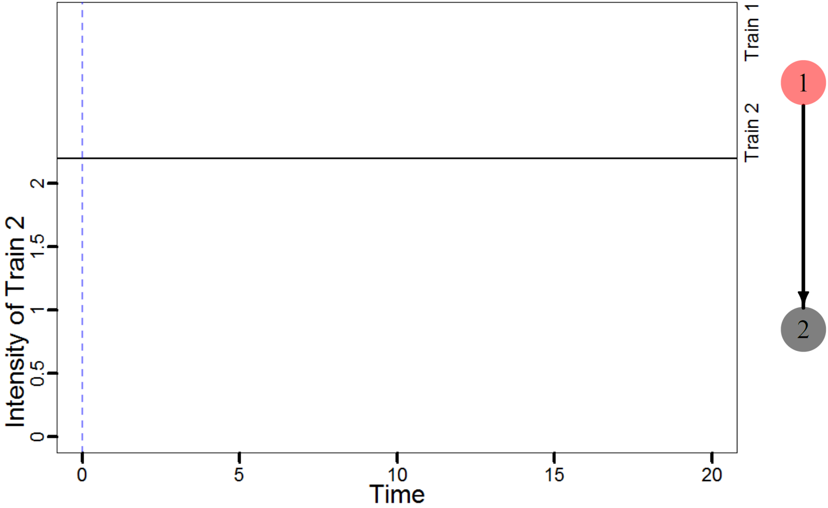
Hawkes (1971)

The Hawkes Process



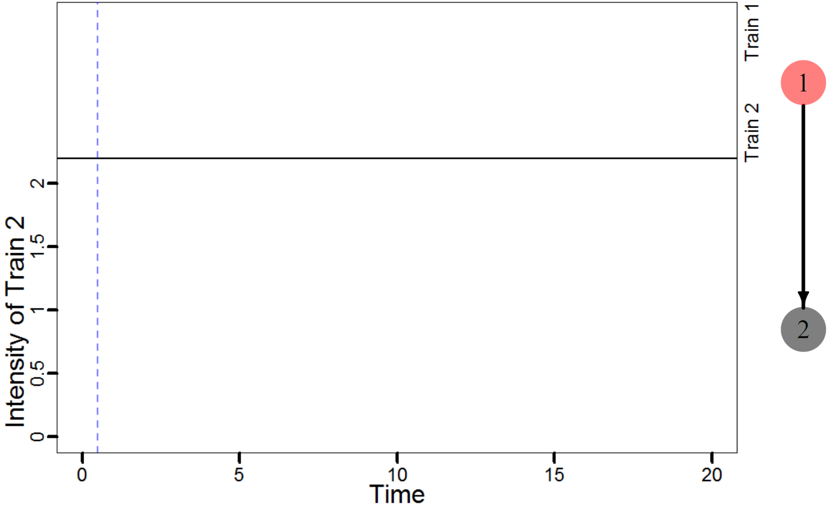
Hawkes (1971)

The Hawkes Process



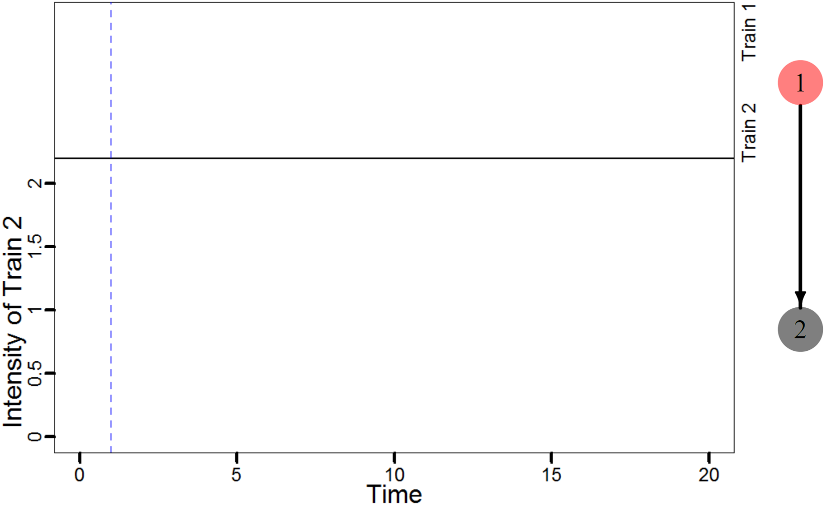
Hawkes (1971)

The Hawkes Process



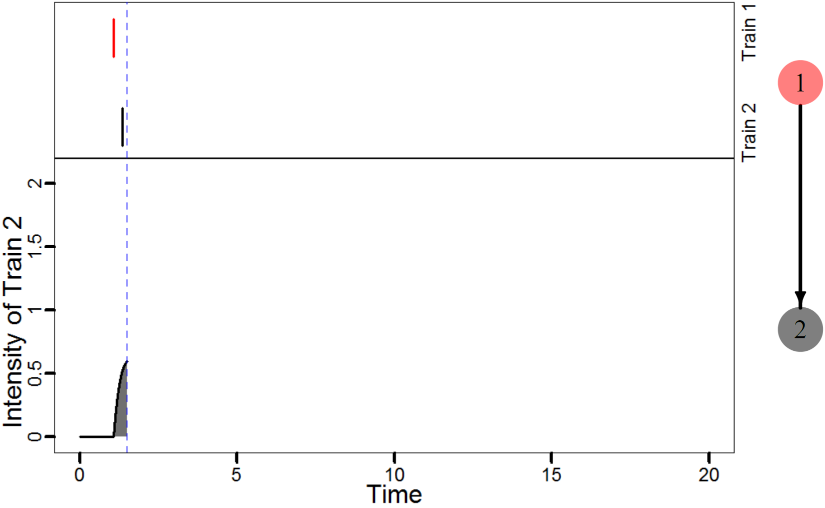
Hawkes (1971)

The Hawkes Process



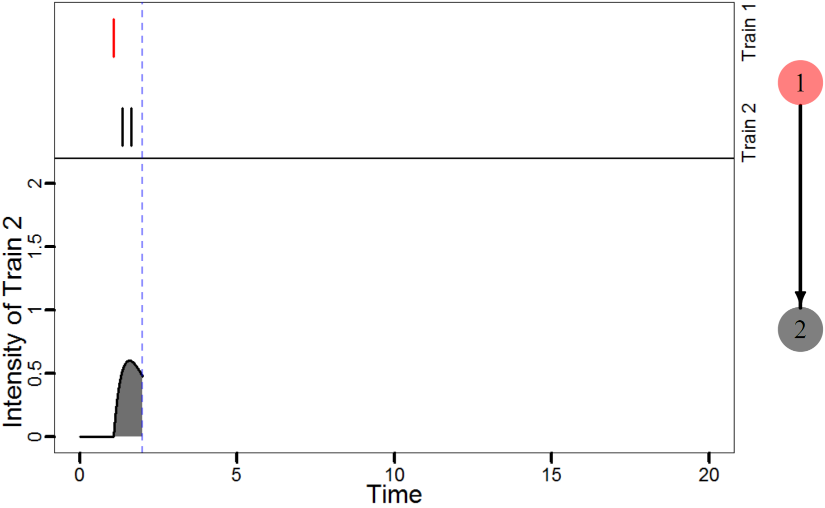
Hawkes (1971)

The Hawkes Process



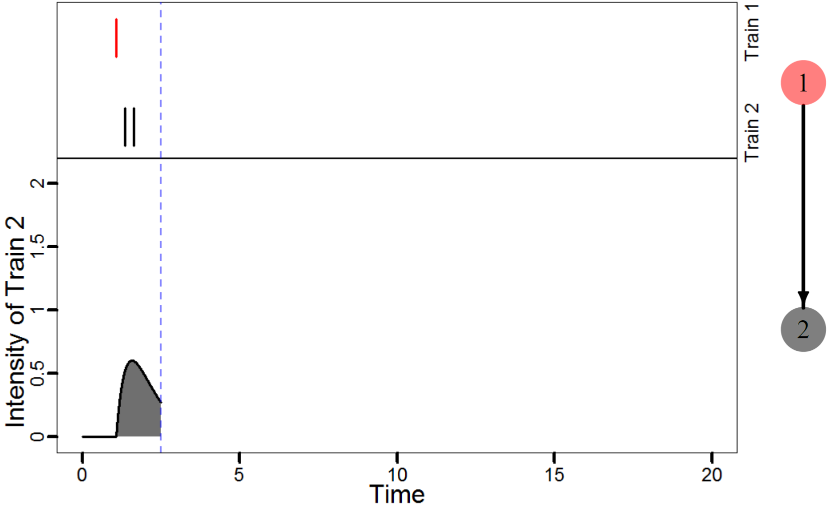
Hawkes (1971)

The Hawkes Process



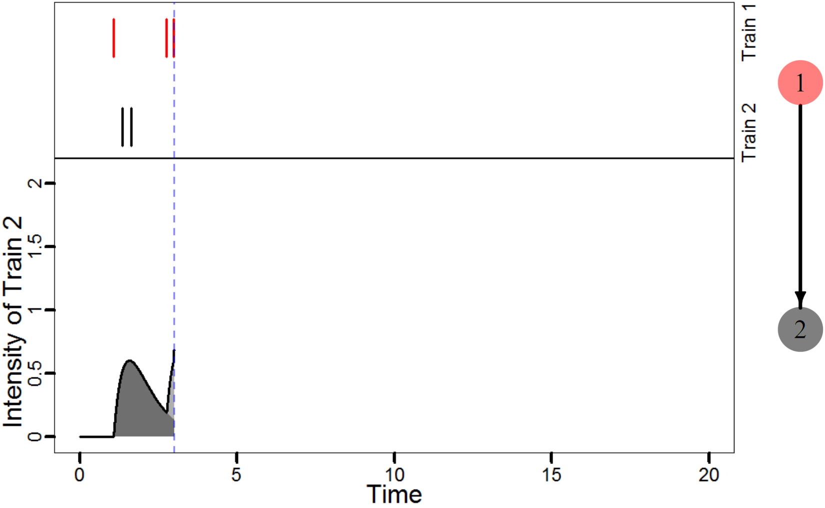
Hawkes (1971)

The Hawkes Process



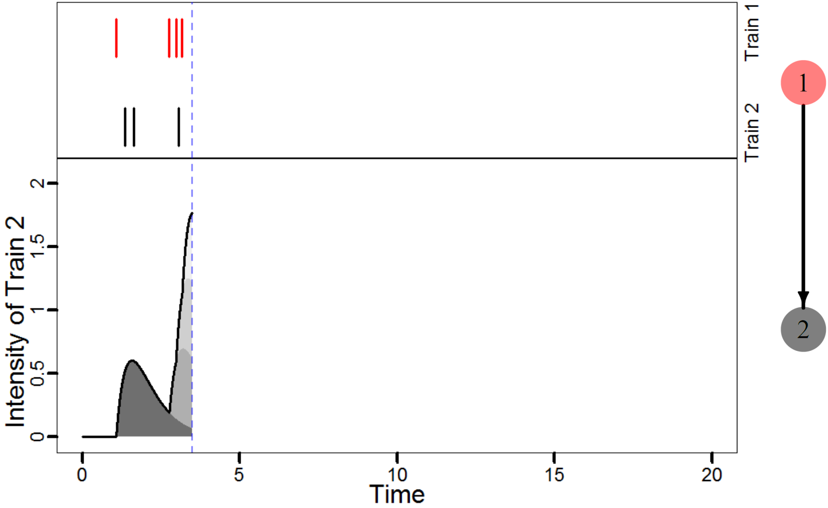
Hawkes (1971)

The Hawkes Process



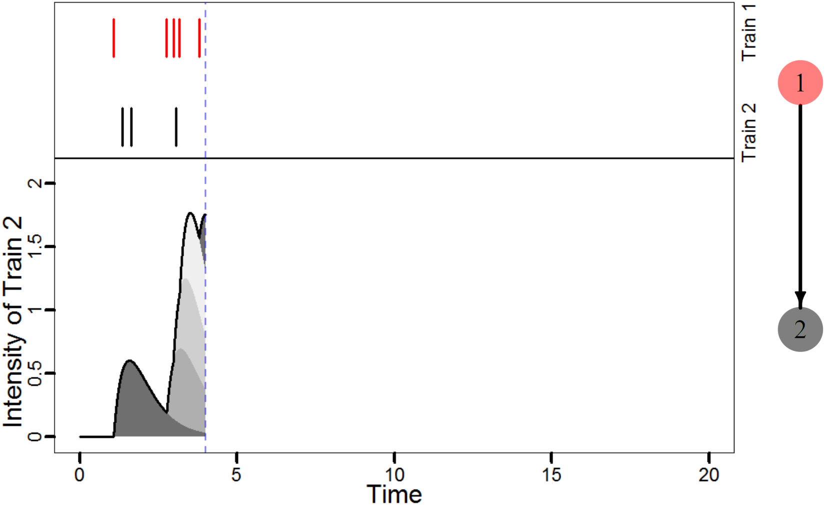
Hawkes (1971)

The Hawkes Process



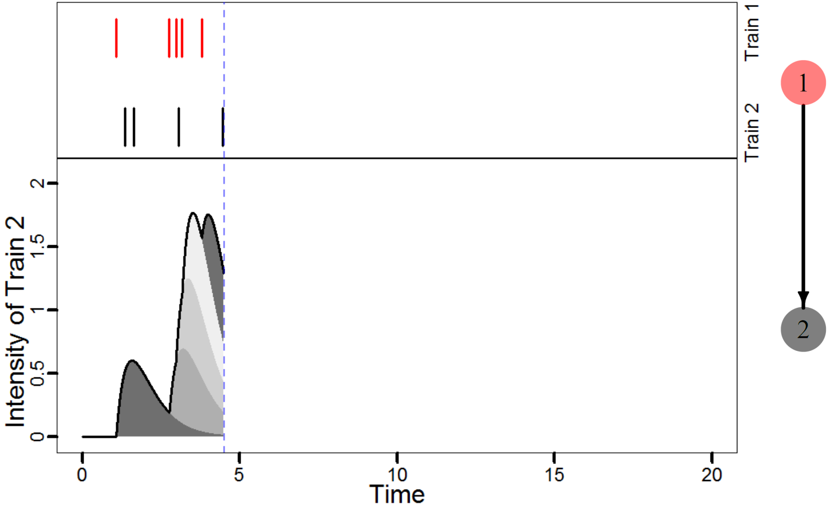
Hawkes (1971)

The Hawkes Process



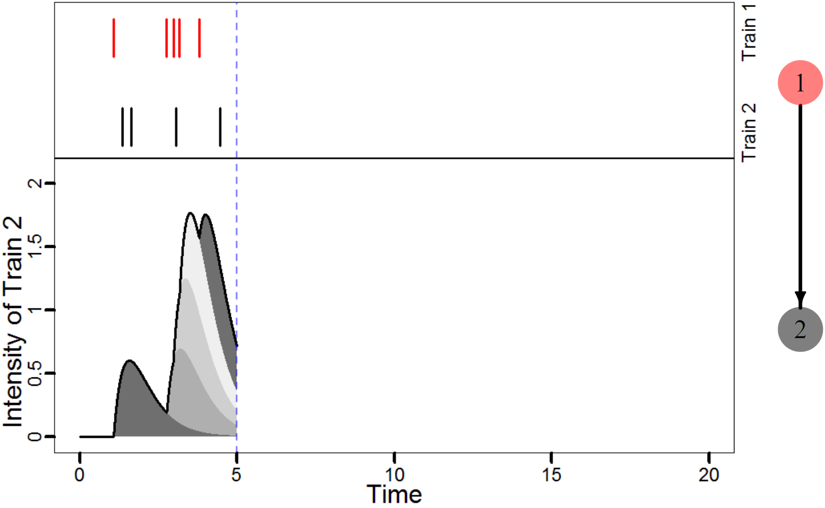
Hawkes (1971)

The Hawkes Process



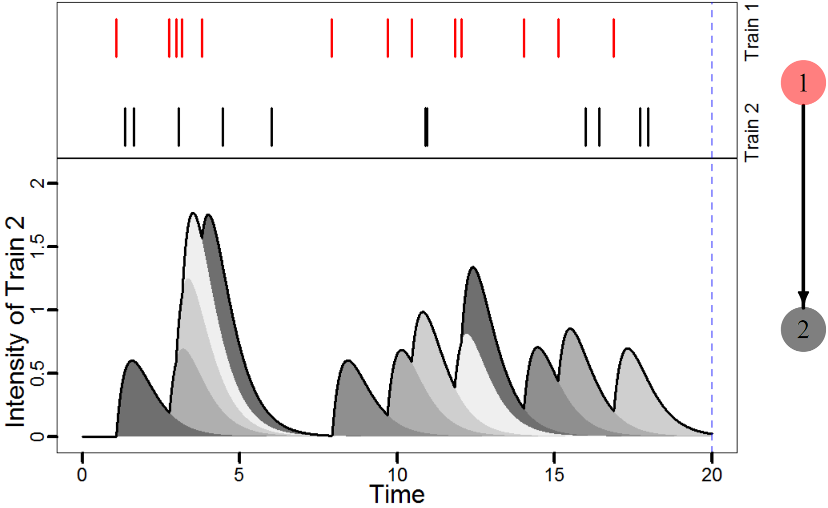
Hawkes (1971)

The Hawkes Process



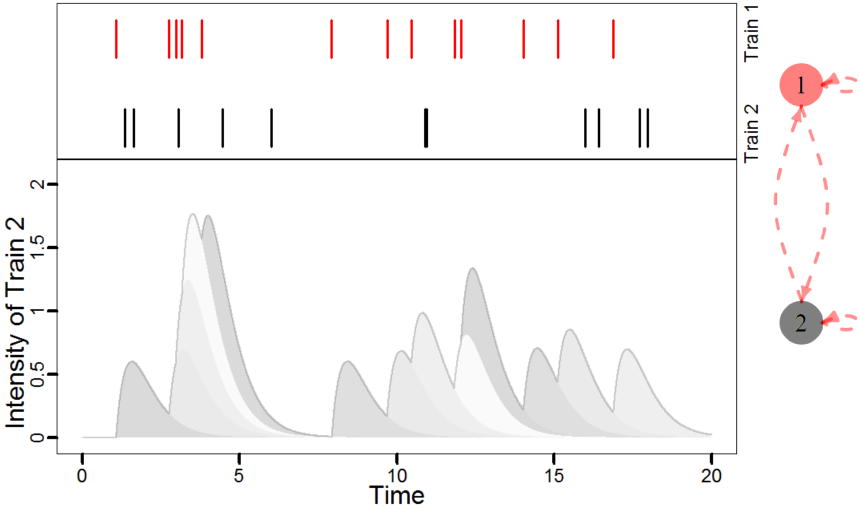
Hawkes (1971)

The Hawkes Process



Hawkes (1971)

Goal



Train 1
Train 2
Intensity of Train 2
0
0.5
1
1.5
2
0
5
10
15
20
Time
1
2

The Hawkes Process

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

- $\lambda_j(\cdot) : \mathbb{R}^+ \to \mathbb{R}$: intensity function
- $\mu_j \in \mathbb{R}$: background intensity
- $\omega_{j,k}(\cdot) : \mathbb{R}^+ \to \mathbb{R}$: transfer function
- $t_{k,i} \in \mathbb{R}^+$: time at which the kth neuron has its ith spike

Hawkes (1971)

Graph Corresponding to the Hawkes Process

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

Graph Corresponding to the Hawkes Process

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

$\omega_{1,2}(t) \neq 0$

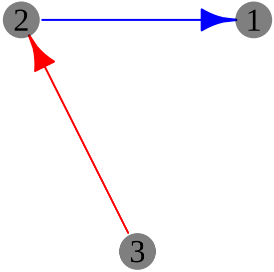
$\omega_{2,3}(t) \neq 0$

Graph Corresponding to the Hawkes Process

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

$\omega_{1,2}(t) \neq 0$

$\omega_{2,3}(t) \neq 0$



Challenges in Fitting the Model, Part I

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

Challenges in Fitting the Model, Part I

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

Challenge: The transfer function $\omega_{j,k}(\cdot)$ is unknown.

Challenges in Fitting the Model, Part I

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

Challenge: The transfer function $\omega_{j,k}(\cdot)$ is unknown.

Solution: Approximate with basis functions, $\psi_1(\cdot), \ldots, \psi_M(\cdot)$:

$$\lambda_j(t) \approx \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \left[\psi(t - t_{k,i})\right]^{\mathrm{T}} \beta_{jk}$$

Challenges in Fitting the Model, Part II

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \le t} \omega_{j,k}(t - t_{k,i})$$

Challenges in Fitting the Model, Part II

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

Challenge: Need to estimate p^2 transfer functions, where p is large.

Challenges in Fitting the Model, Part II

$$\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})$$

Challenge: Need to estimate p^2 transfer functions, where p is large.

Solution: Group lasso to induce sparsity in transfer functions.

Our Proposal: Neighborhood Selection Approach

Related Work: Meinshausen and Bühlmann (2006); Zhou et al. (2013a,b); Bacry et al. (2015); Hansen et al. (2015)

Our Proposal: Neighborhood Selection Approach

Step 1: For $j = 1, \ldots, p$, find $\hat{\beta}_{j1}, \ldots, \hat{\beta}_{jp} \in \mathbb{R}^M$ that minimize

$$L_j(\beta_{j1}, \ldots, \beta_{jp}) + \lambda \sum_{k=1}^{p} \|\psi^{\mathrm{T}} \beta_{j,k}\|_2.$$

Related Work: Meinshausen and Bühlmann (2006); Zhou et al. (2013a,b); Bacry et al. (2015); Hansen et al. (2015)

Our Proposal: Neighborhood Selection Approach

Step 1: For $j = 1, \ldots, p$, find $\hat{\beta}_{j1}, \ldots, \hat{\beta}_{jp} \in \mathbb{R}^M$ that minimize

$$L_j(\beta_{j1}, \ldots, \beta_{jp}) + \lambda \sum_{k=1}^{p} \|\psi^{\mathrm{T}} \beta_{j,k}\|_2.$$

Step 2: The graph estimate is $\hat{\mathcal{E}} = \left\{ (j, k) : \hat{\beta}_{jk} \neq 0 \right\}$.

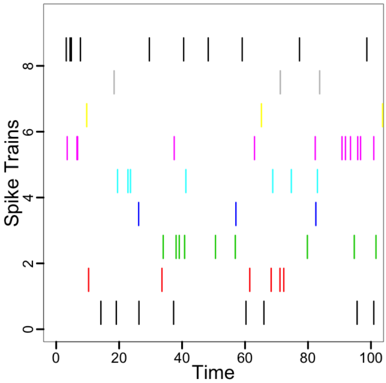
Related Work: Meinshausen and Bühlmann (2006); Zhou et al. (2013a,b); Bacry et al. (2015); Hansen et al. (2015)

Theoretical Results

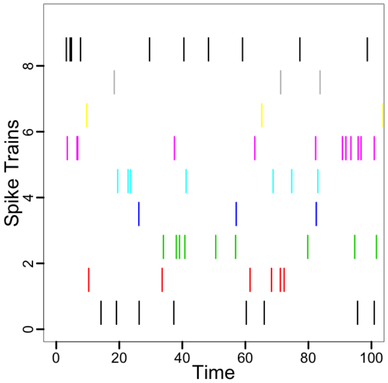
We establish model selection consistency in high dimensions; i.e. the parent set of each neuron is correctly estimated.

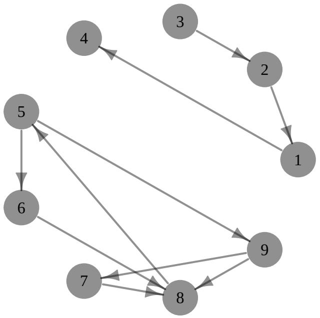
The End Result

The End Result



The End Result





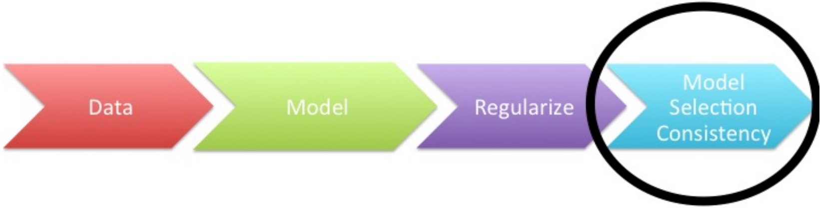
Summary of Pipeline

Summary of Pipeline

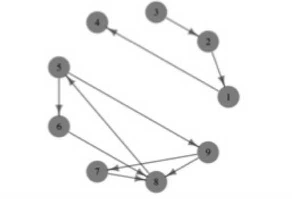
Summary of Pipeline

Summary of Pipeline

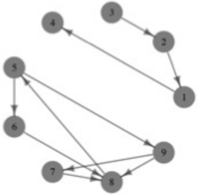
Summary of Pipeline

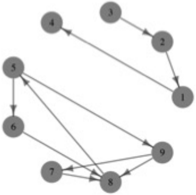


Model Selection Consistency

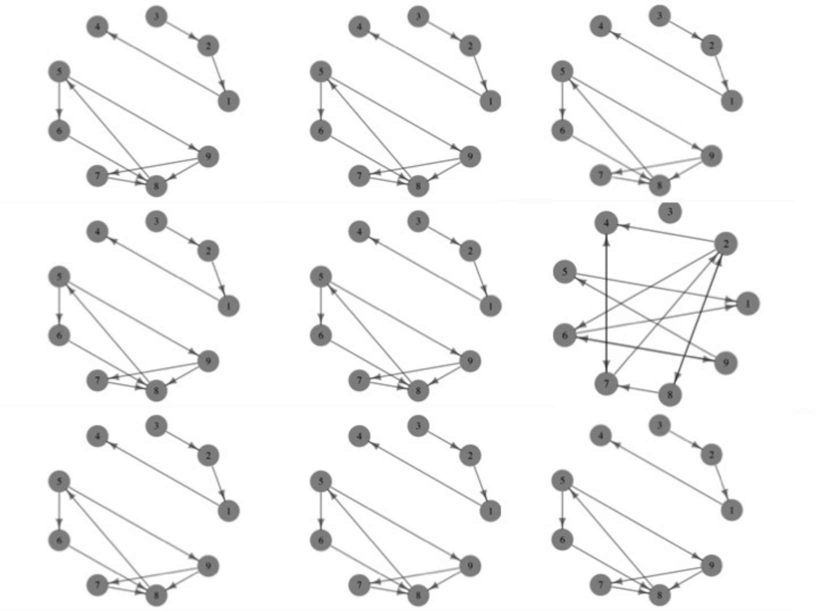


Model Selection Consistency

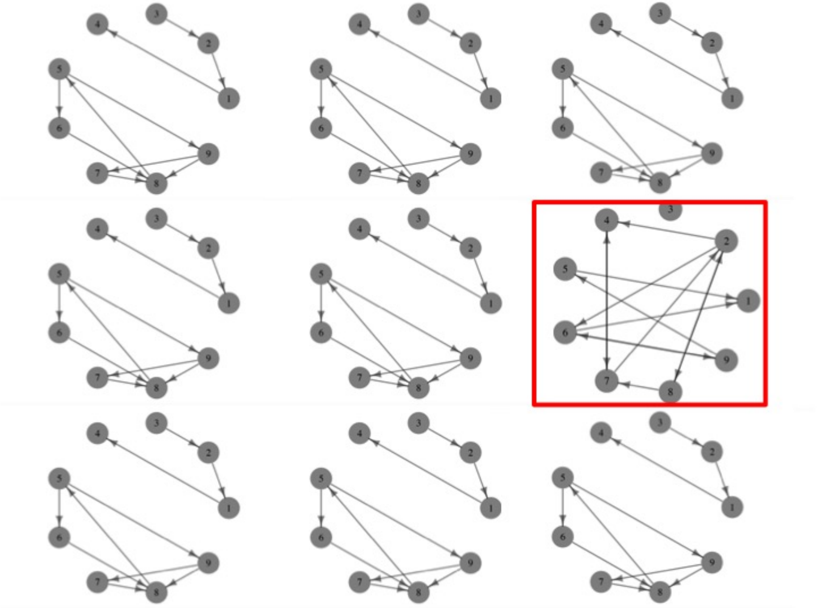




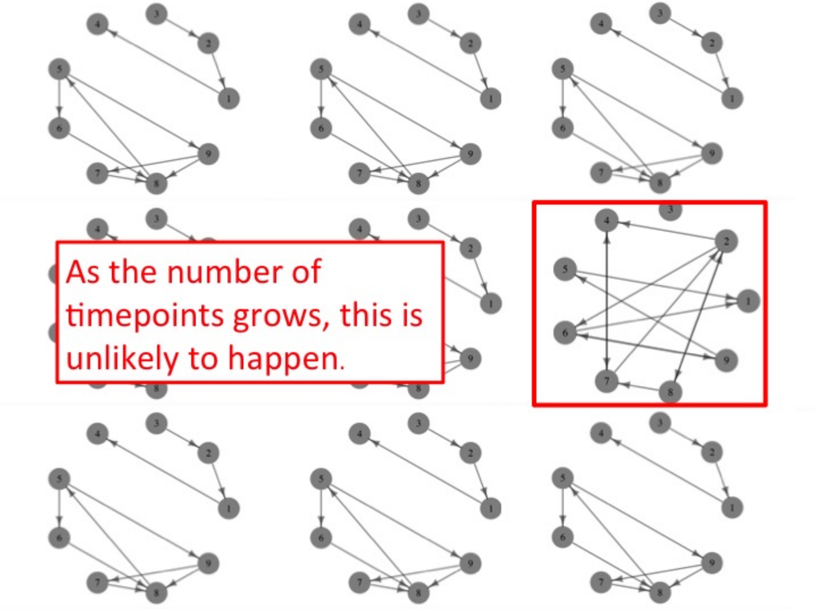
Model Selection Consistency



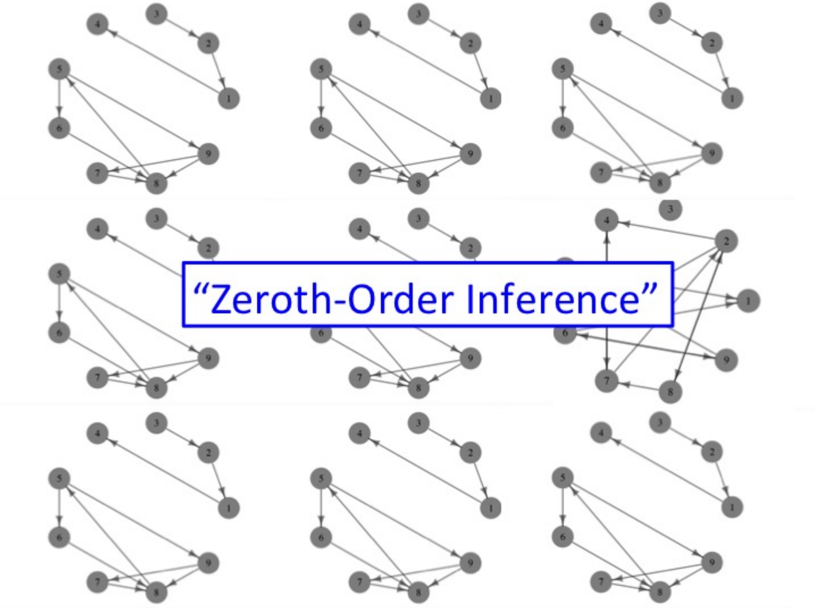
Model Selection Consistency



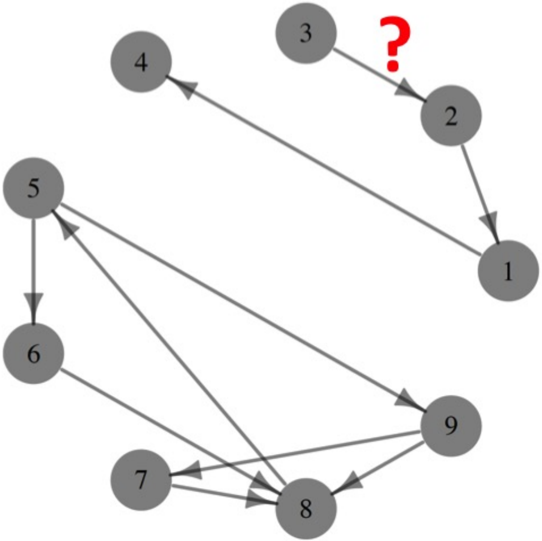
Model Selection Consistency



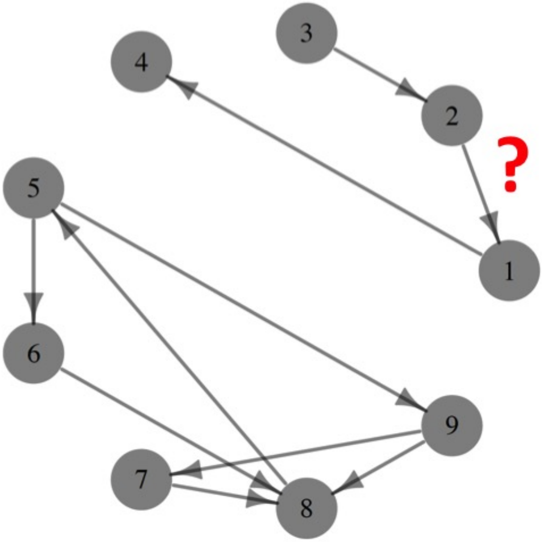
Model Selection Consistency



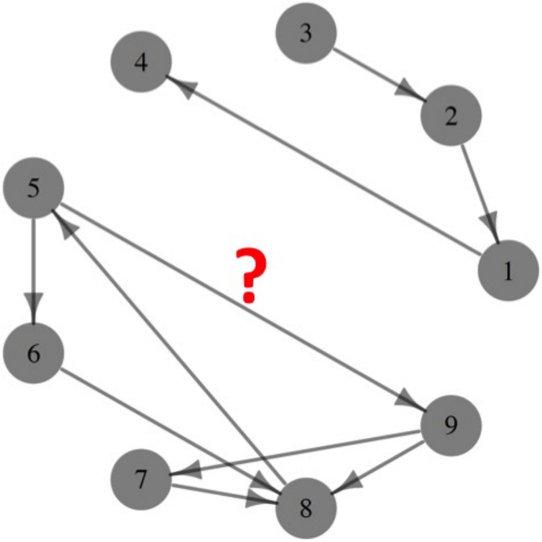
What Does First-Order Inference Look Like?



What Does First-Order Inference Look Like?



What Does First-Order Inference Look Like?



What Does First-Order Inference Look Like?

Early work in this direction: papers by J. Taylor, R. & R. Tibshirani, C.-H. Zhang, M. Buhlmann, A. Montanari, S. van de Geer, and many others

What Does First-Order Inference Look Like?

- P-value associated with each edge?

Early work in this direction: papers by J. Taylor, R. & R. Tibshirani, C.-H. Zhang, M. Buhlmann, A. Montanari, S. van de Geer, and many others

What Does First-Order Inference Look Like?

- P-value associated with each edge?
- False discovery rate associated with the estimated edge set?

Early work in this direction: papers by J. Taylor, R. & R. Tibshirani, C.-H. Zhang, M. Buhlmann, A. Montanari, S. van de Geer, and many others

What Does First-Order Inference Look Like?

- P-value associated with each edge?
- False discovery rate associated with the estimated edge set?
- Posterior distribution?

Early work in this direction: papers by J. Taylor, R. & R. Tibshirani, C.-H. Zhang, M. Buhlmann, A. Montanari, S. van de Geer, and many others

What Does First-Order Inference Look Like?

- P-value associated with each edge?
- False discovery rate associated with the estimated edge set?
- Posterior distribution?

Model certainly does not hold!

Early work in this direction: papers by J. Taylor, R. & R. Tibshirani, C.-H. Zhang, M. Buhlmann, A. Montanari, S. van de Geer, and many others

Summary

Summary

- Learn graph structure from temporal data.

Summary

- Learn graph structure from temporal data.
- Different data, different models.

Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:

Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:
  - Do not assume functional form: use basis expansions.

Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:
  - Do not assume functional form: use basis expansions.
  - Estimate a sparse graph using group lasso penalties.

Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:
  - Do not assume functional form: use basis expansions.
  - Estimate a sparse graph using group lasso penalties.
  - Establish that the estimated graph is correct w.h.p.

What is Missing?

What is Missing?

- Model and assumptions certainly do not hold. Now what?

What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.

What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.
- Can I get a p-value for each edge, or a false discovery rate?

What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.
- Can I get a p-value for each edge, or a false discovery rate?
- Do I really believe the estimated graph?

What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.
- Can I get a p-value for each edge, or a false discovery rate?
- Do I really believe the estimated graph?
  - Next steps for a biological collaborator?
  - No gold standard.

References

Bacry, E., S. Gaïffas, and J.-F. Muzy (2015). A generalization error bound for sparse and low-rank multivariate Hawkes processes. *arXiv preprint arXiv:1501.00725*.

Dattner, I. and C. A. Klaassen (2013). Estimation in systems of ordinary differential equations linear in the parameters. *arXiv preprint arXiv:1305.4126*.

Hansen, N. R., P. Reynaud-Bouret, and V. Rivoirard (2015, 02). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli 21*(1), 83–143.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*(1), 83–90.

Henderson, J. and G. Michailidis (2014). Network reconstruction using nonparametric additive ODE models. *PLoS ONE 9*(4), e94003.

Loh, P.-L. and M. J. Wainwright (2012). High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity. *Ann. Statist. 40*(3), 1637–1664.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist. 34*(3), 1436–1462.

Pillow, J. W., J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature 454*(7207), 995–999.

Ravikumar, P. K., J. Lafferty, H. Liu, and L. Wasserman (2009). Sparse additive models. *J. R. Stat. Soc. Ser. B Stat. Methodol. 71*(5), 1009–1030.

Schaffter, T., D. Marbach, and D. Floreano (2011). GeneNetWeaver: *in silico* benchmark generation and performance profiling of network inference methods. *27*(16), 2263–2270.

Simon, N. and R. J. Tibshirani (2012). Standardization and the group lasso penalty. *Statist. Sinica 22*(3), 983–1001.

Wu, H., T. Lu, H. Xue, and H. Liang (2014). Sparse additive ordinary differential equations for dynamic gene regulatory network modeling. *J. Amer. Statist. Assoc. 109*(506), 700–716.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol. 68*(1), 49–67.

Zhou, K., H. Zha, and L. Song (2013a). Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 641–649.

Zhou, K., H. Zha, and L. Song (2013b). Learning triggering kernels for multi-dimensional Hawkes processes. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pp. 1301–1309.