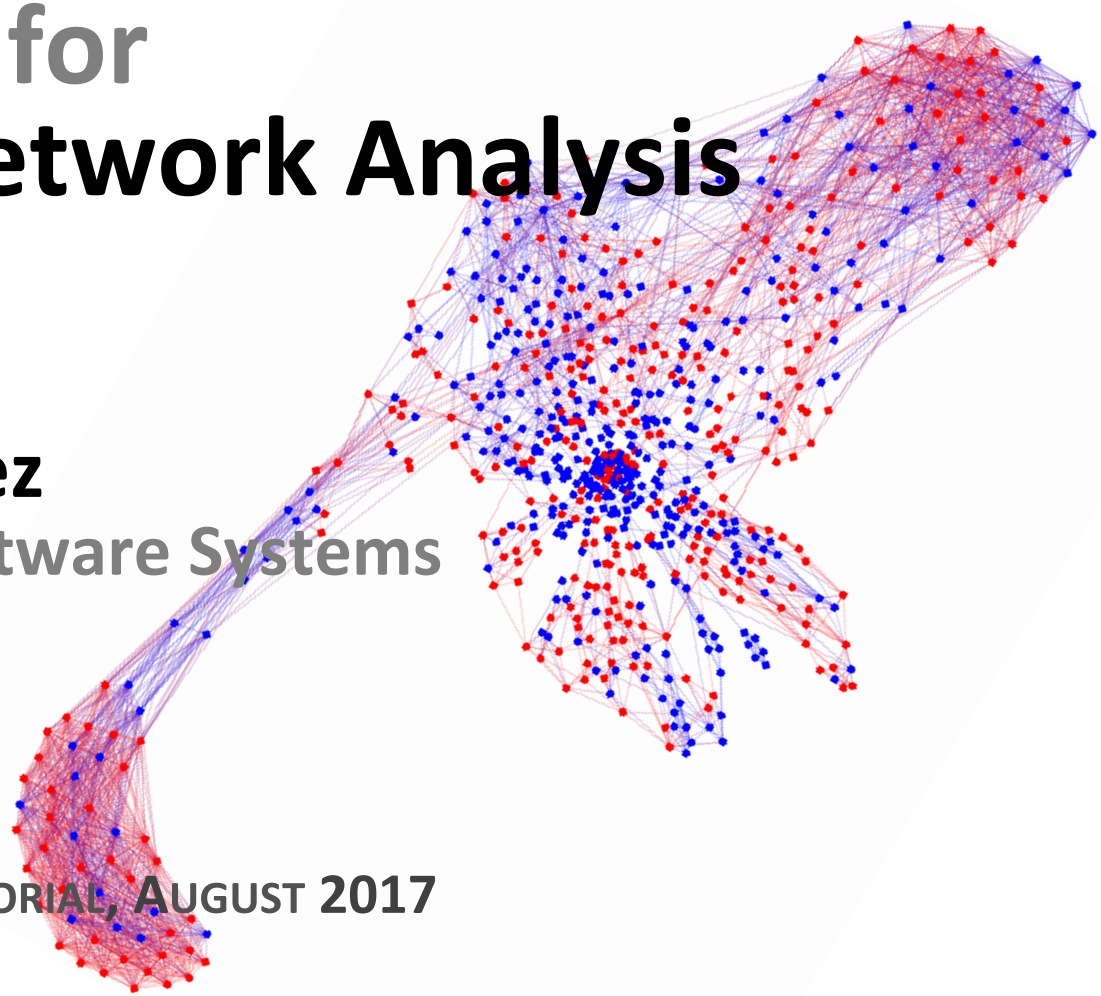
Machine learning for
Dynamic Social Network Analysis



Manuel Gomez Rodriguez

Max Planck Institute for Software Systems

IJCAI Tutorial, August 2017

Many discrete *events* in continuous time

Qmee, 2013

Variety of processes behind these events

Events are (noisy) observations of a variety of complex dynamic processes…

News spread in Twitter

Product reviews and sales in Amazon

A user gains recognition in Quora

Video becomes viral in Youtube

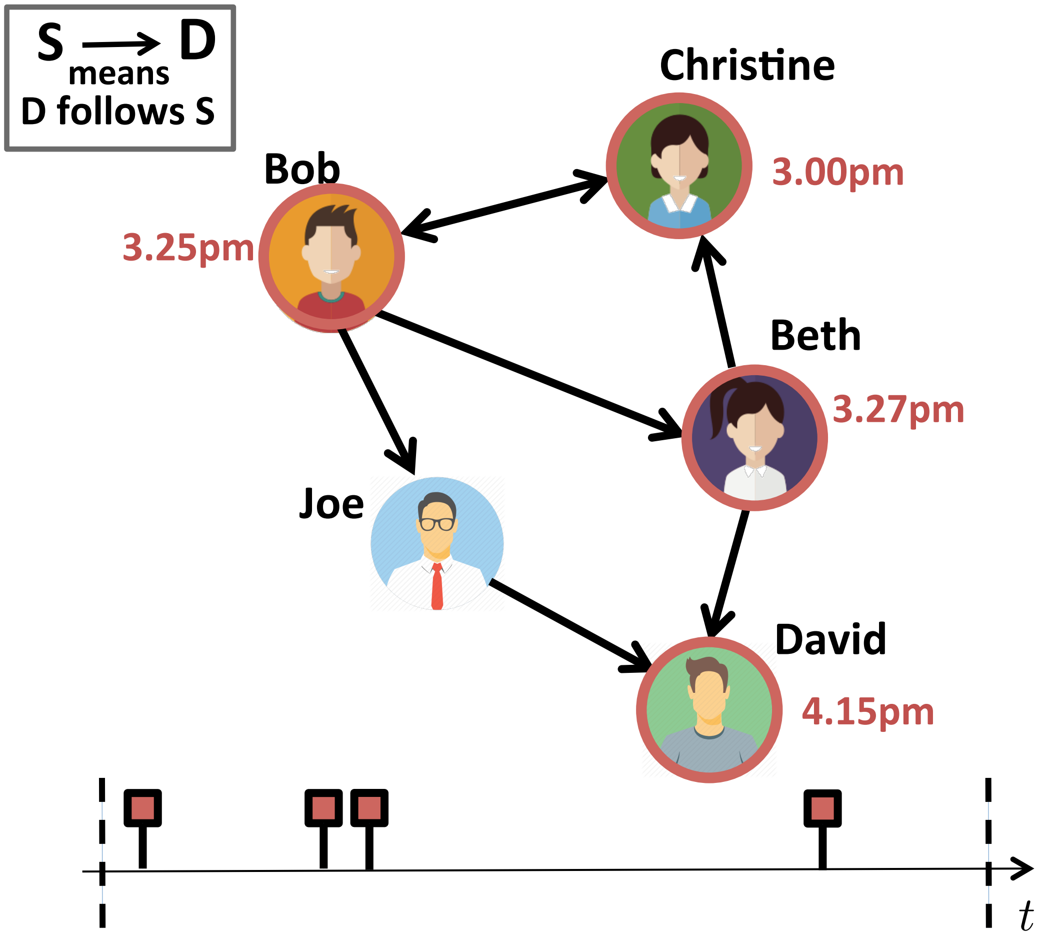
Article creation in Wikipedia

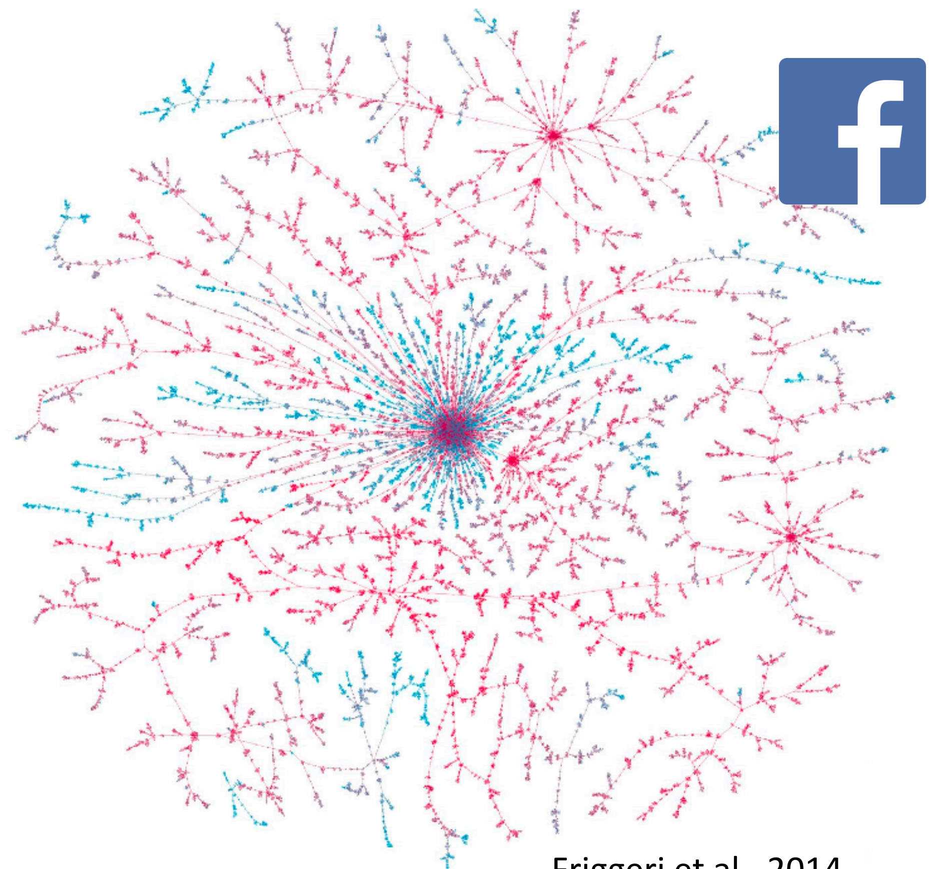
FAST

SLOW

…in a wide range of temporal scales.

Example I: Idea adoption/viral marketing





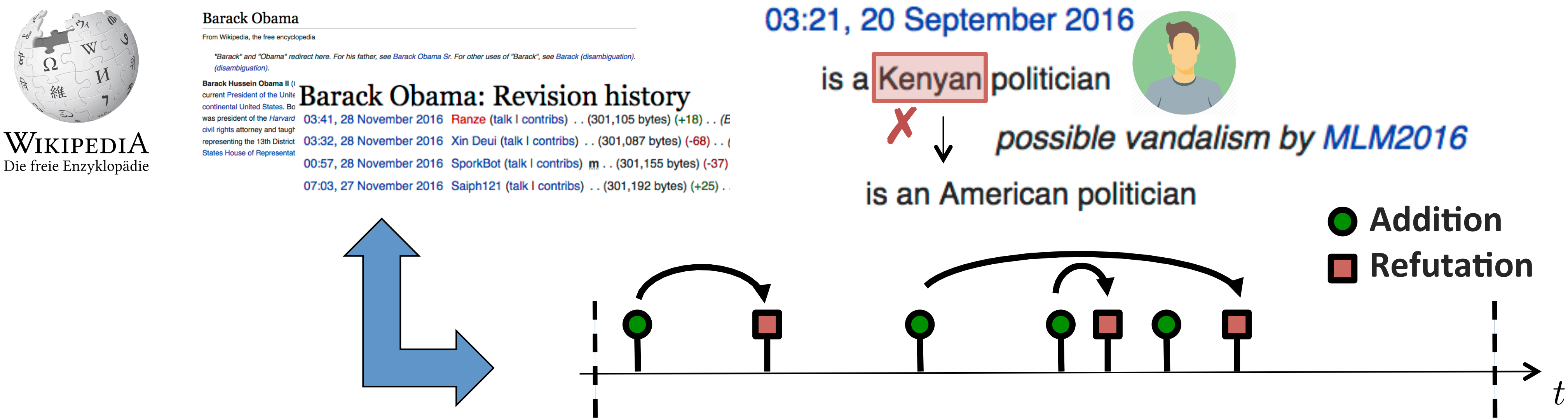
Friggeri et al., 2014

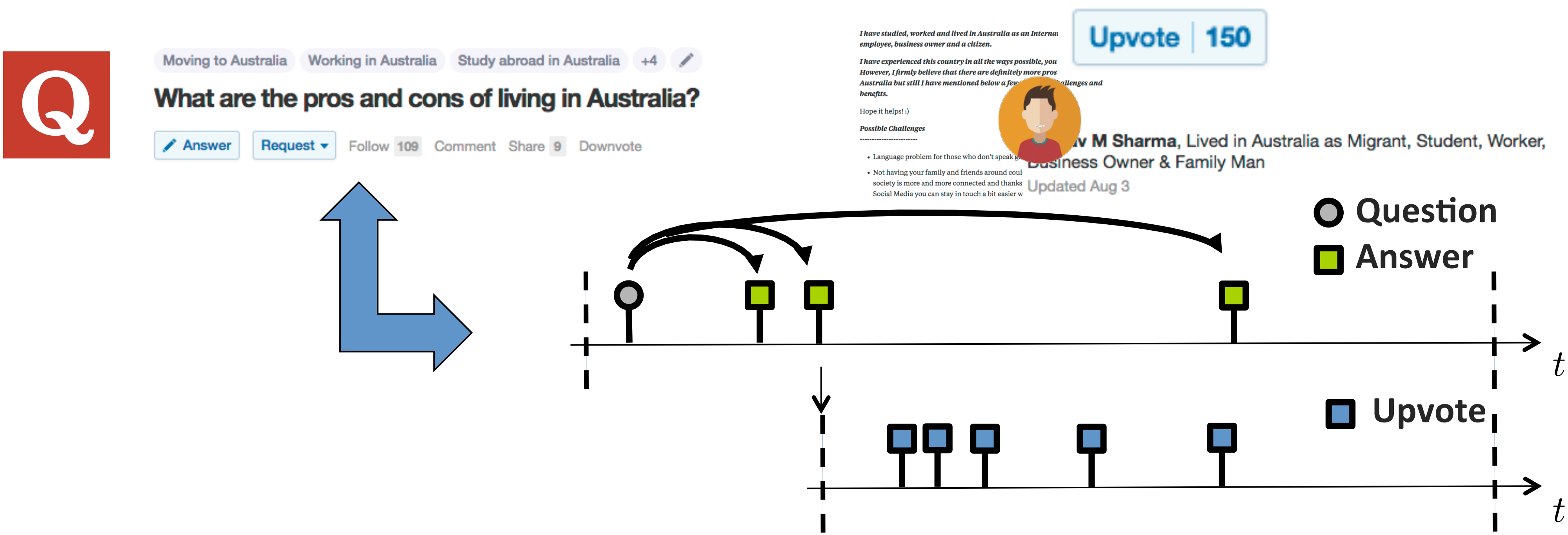
They can have an impact in the off-line world

theguardian

Click and elect: how fake news helped Donald Trump win a real election

Example II: Information creation & curation

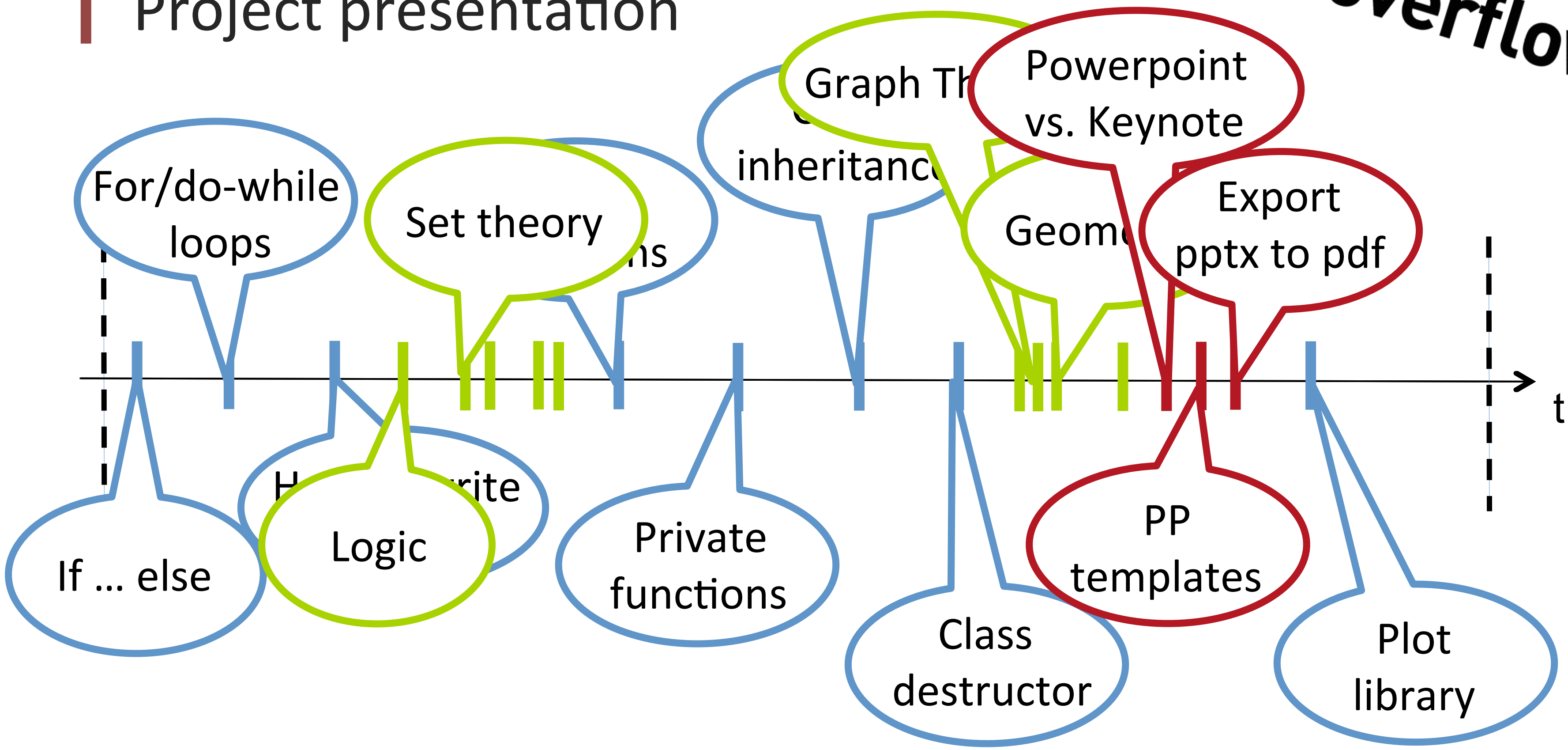




Example III: Learning trajectories

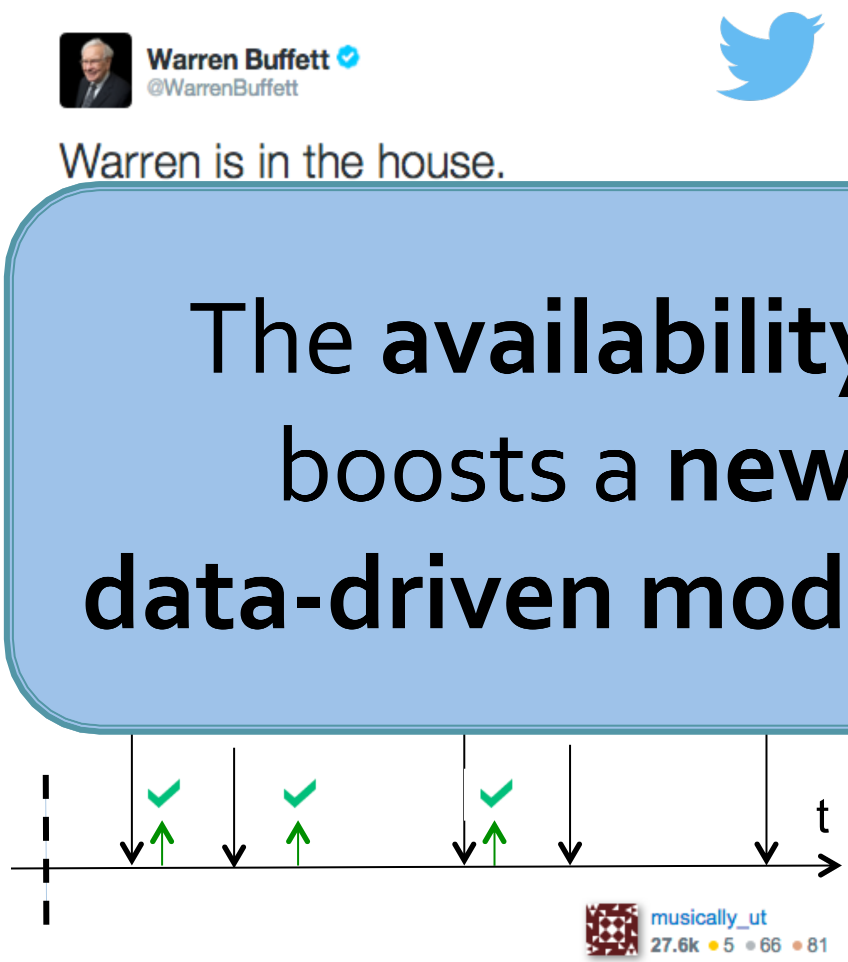
1st year computer science student

- Introduction to programming
- Discrete math
- Project presentation



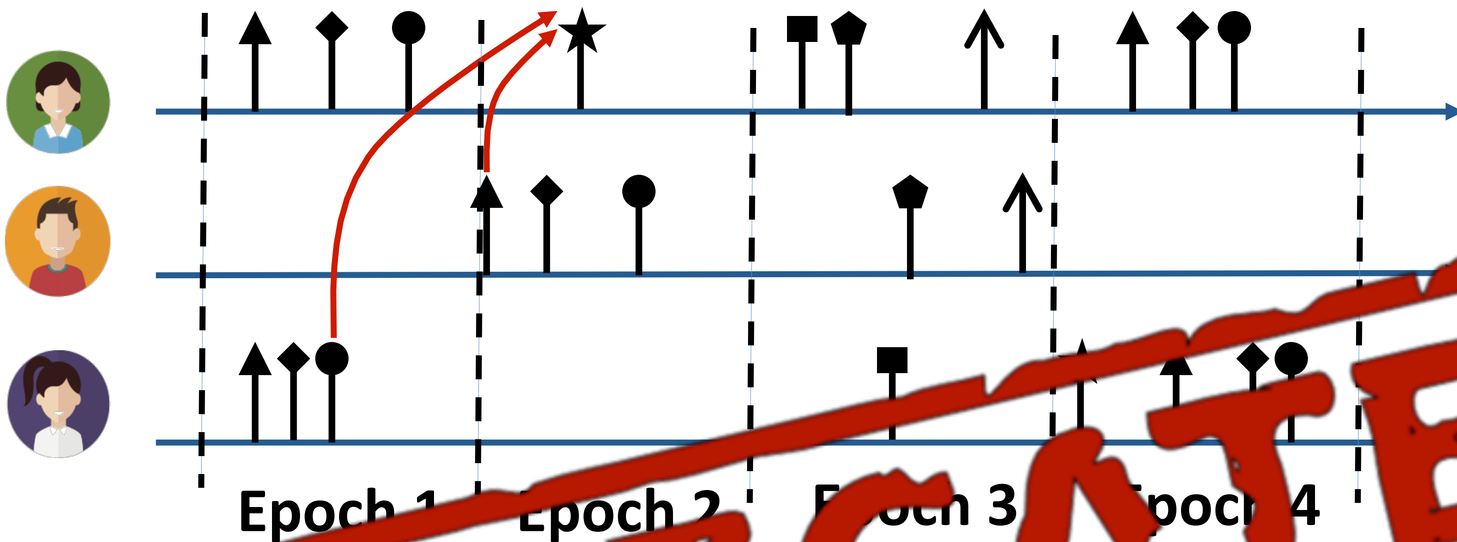
Detailed *event traces*

DETAILED TRACES OF ACTIVITY



The **availability of event traces** boosts a **new generation** of **data-driven models and algorithms**

Previously: discrete-time models & algorithms



DEPRECATED

Discrete-time models artificially introduce epochs:

1. How long is each epoch? Data is very heterogeneous.

2. How to aggregate events within an epoch?

3. What if no event within an epoch?

4. Time is treated as index or conditioning variable, not easy to deal with time-related queries.

Outline of the Seminar

Representation: Temporal Point processes

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

Next

Applications: Models

1. Information propagation
2. Information reliability
3. Knowledge acquisition

Applications: Control

1. Activity shaping
2. When-to-post

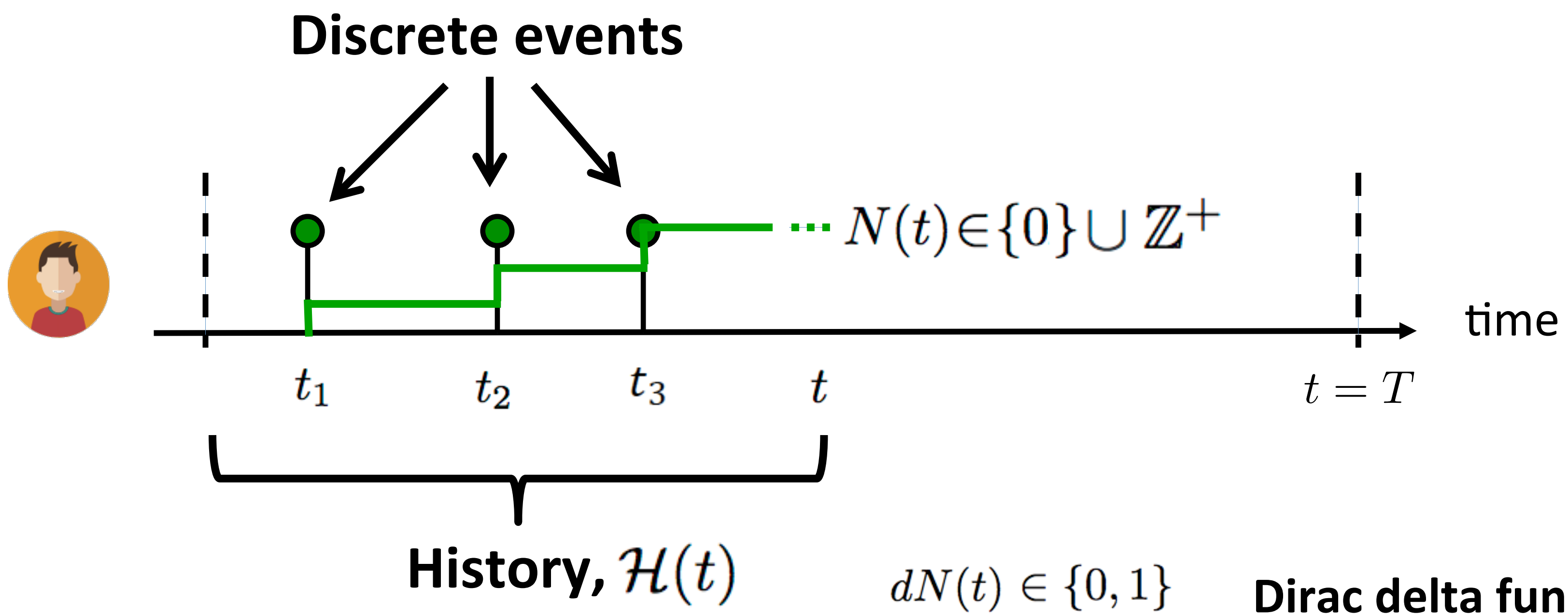
Slides/references: learning.mpi-sws.org/ijcai-2017-tutorial

Representation:
Temporal Point Processes

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

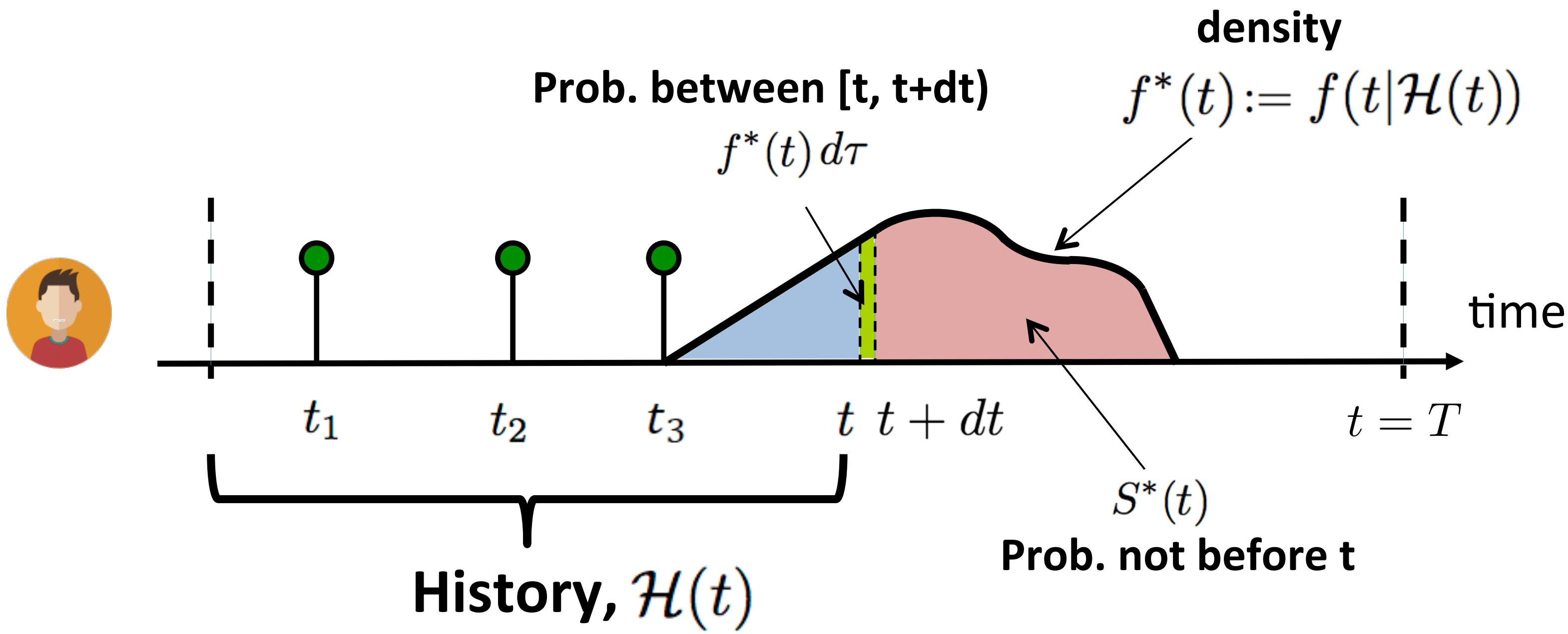
Temporal point processes

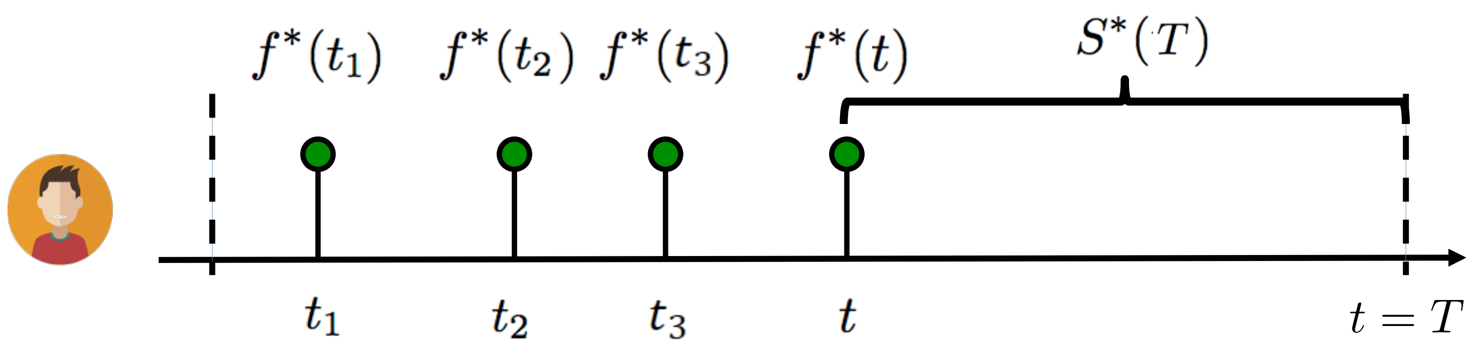
Temporal point process:
A random process whose realization consists of **discrete events localized in time**



Formally: $N(t) = \int_0^t dN(s) \Rightarrow \overset{dN(t) \in \{0,1\}}{\overset{\downarrow}{dN(t)}} = \sum_{t_i \in \mathcal{H}(t)} \overset{\text{Dirac delta function}}{\overset{\downarrow}{\delta}}(t - t_i)\, dt$

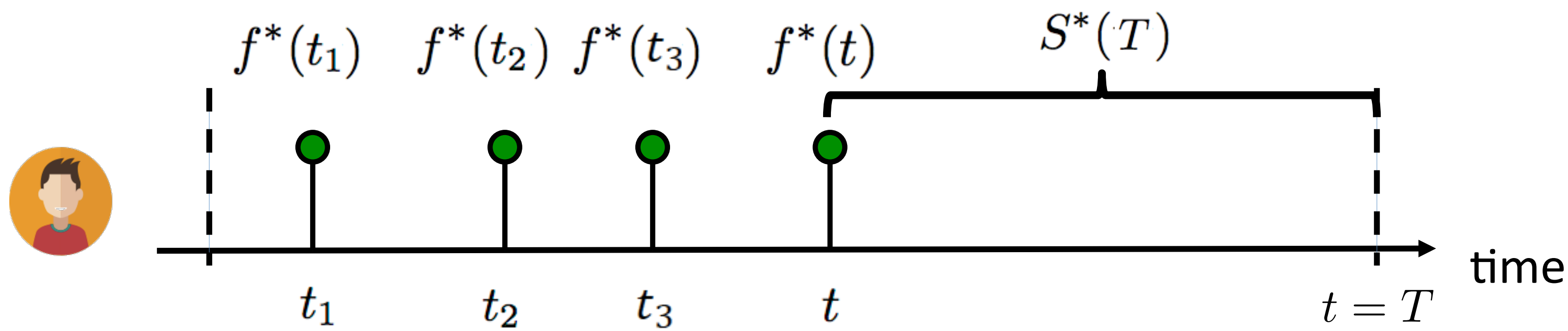
Model time as a random variable





Likelihood of a timeline: $f^*(t_1)\ f^*(t_2)\ f^*(t_3)\ f^*(t)\ S^*(T)$

Problems of density parametrization (I)



$$f^*(t_1) \quad f^*(t_2) \quad f^*(t_3) \quad f^*(t) \quad S^*(T)$$

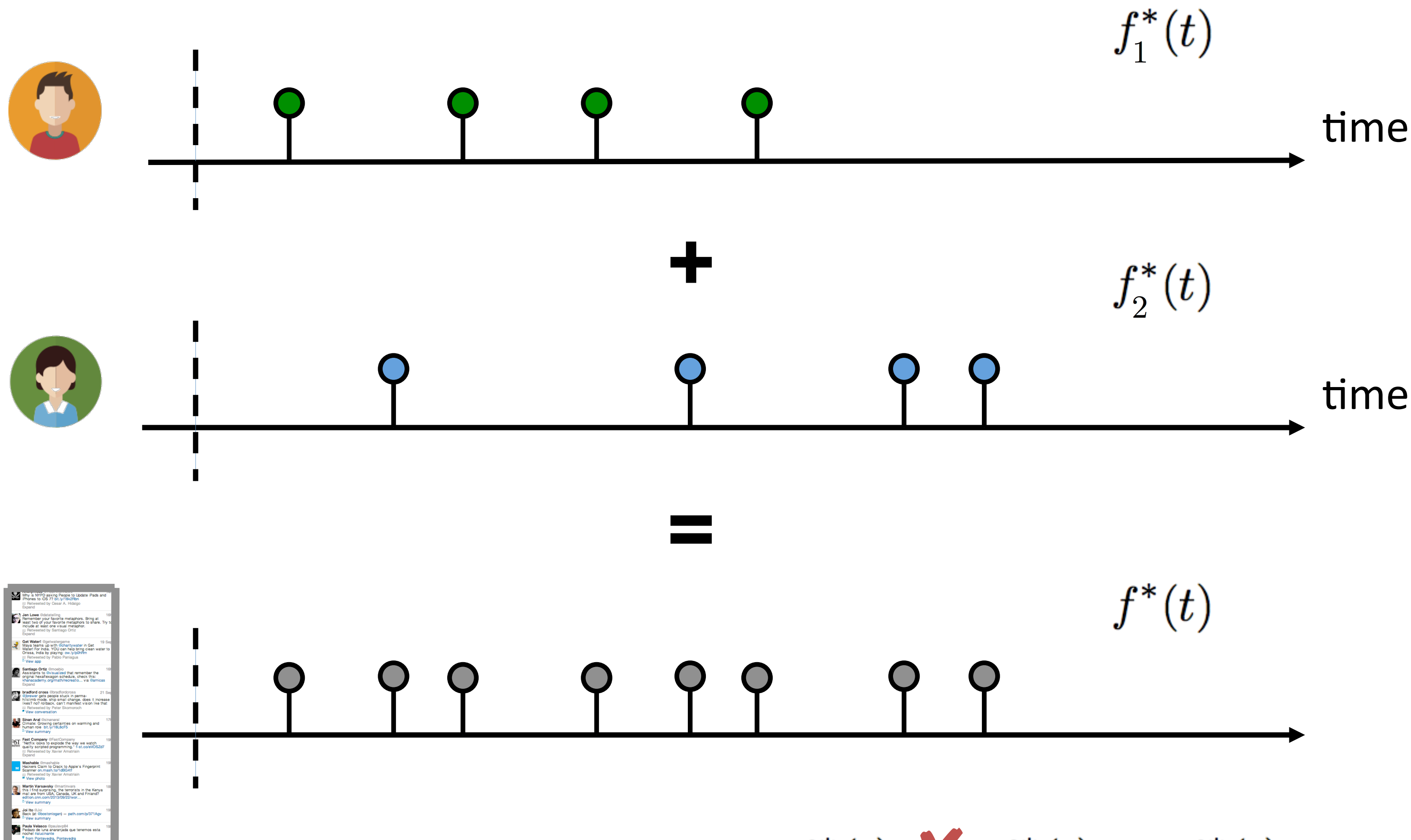
$$\frac{\exp\langle w, \psi^*(t_1)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t_2)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t_3)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t)\rangle}{Z} \quad 1 - \int_t^T \frac{\exp\langle w, \psi^*(\tau)\rangle}{Z} d\tau$$

It is difficult for model design and interpretability:

1. **Densities need to integrate to 1 (i.e., partition function)**
2. **Difficult to combine timelines**

Problems of density parametrization (II)

Difficult to combine timelines:

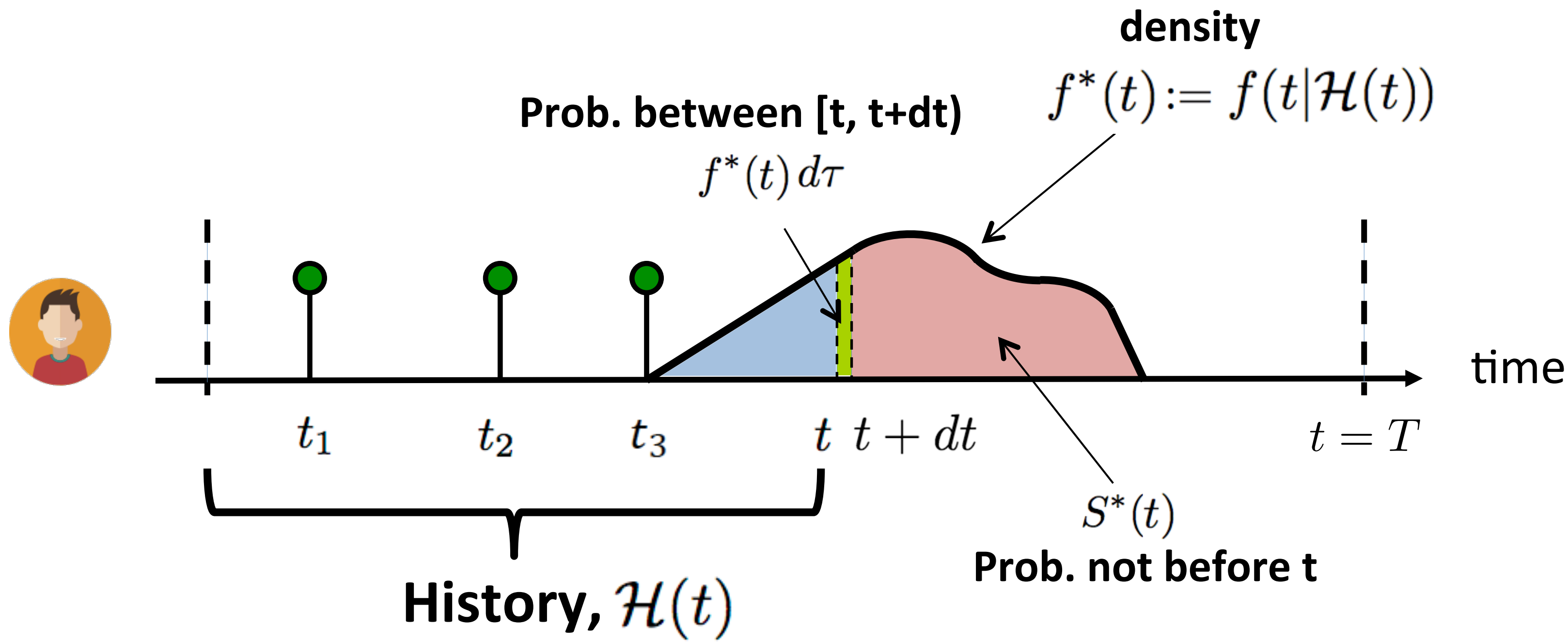


Sum of random processes

$$f^*(t) \cancel{=} f_1^*(t) + f_2^*(t)$$

$$f^*(t) \cancel{=} f_1^*(t) * f_2^*(t)$$

Intensity function



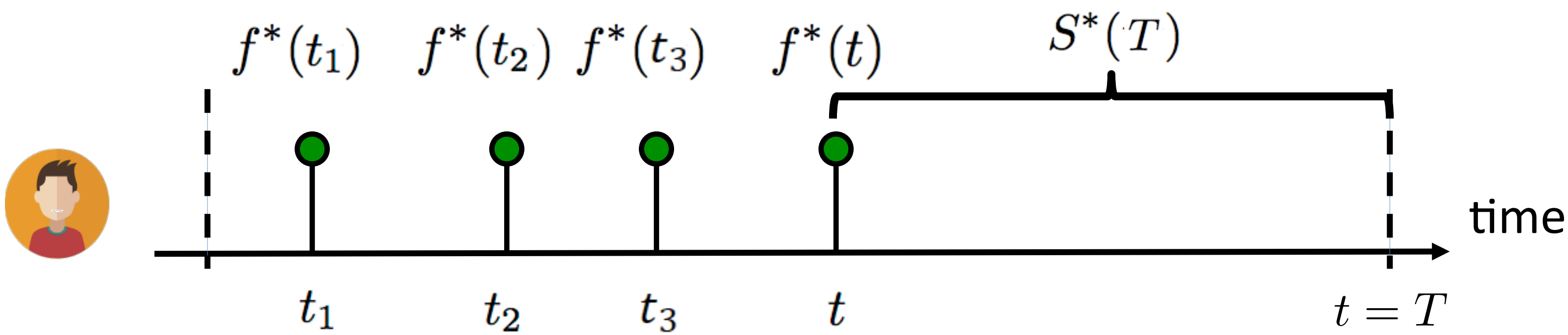
Intensity:

Probability between [t, t+dt) but not before t

$$\lambda^*(t)dt = \frac{f^*(t)dt}{S^*(t)} \geq 0 \quad \Rightarrow \quad \lambda^*(t)dt = \mathbb{E}[dN(t)|\mathcal{H}(t)]$$

Observation: $\lambda^*(t)$ **It is a rate = # of events / unit of time**

Advantages of intensity parametrization (I)



$$\lambda^*(t_1)\ \lambda^*(t_2)\ \lambda^*(t_3)\ \lambda^*(t)\ \exp\left(-\int_0^T \lambda^*(\tau)\, d\tau\right)$$

$\langle w, \phi^*(t_1)\rangle$

$\langle w, \phi^*(t_3)\rangle$

$\langle w, \phi^*(t_2)\rangle$

$\langle w, \phi^*(t)\rangle$

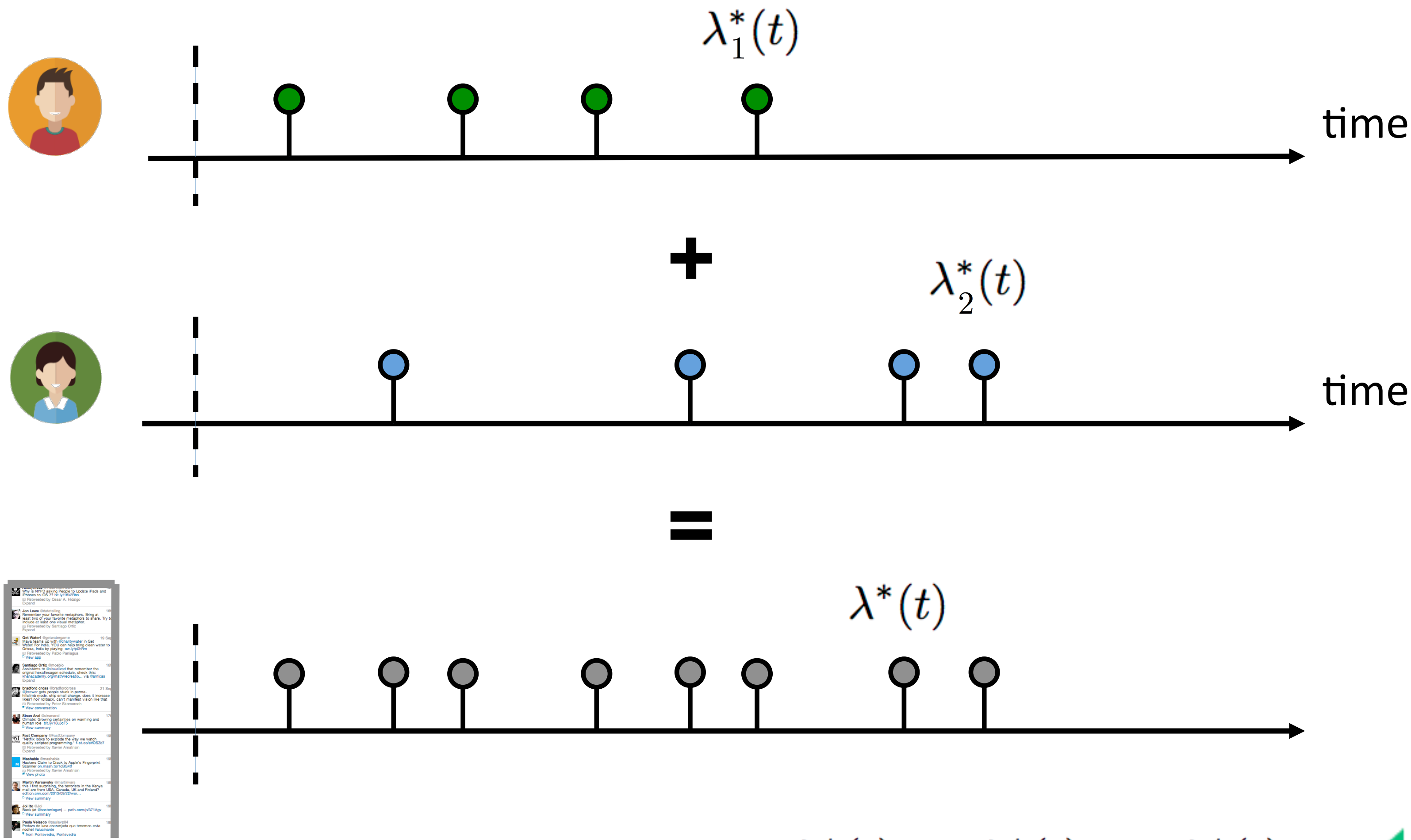
$\exp\left(-\int_0^T \langle w, \phi^*(\tau)\rangle d\tau\right)$

Suitable for model design and interpretable:

1. **Intensities only need to be nonnegative**
2. **Easy to combine timelines**

Advantages of intensity parametrization (II)

Easy to combine timeline:

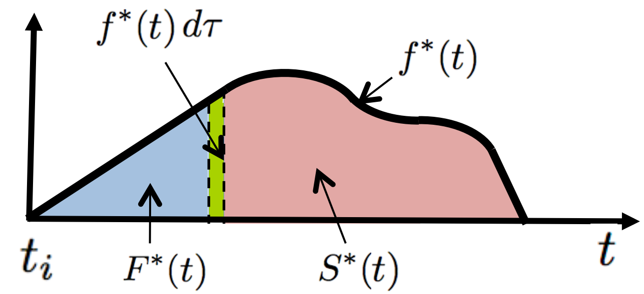


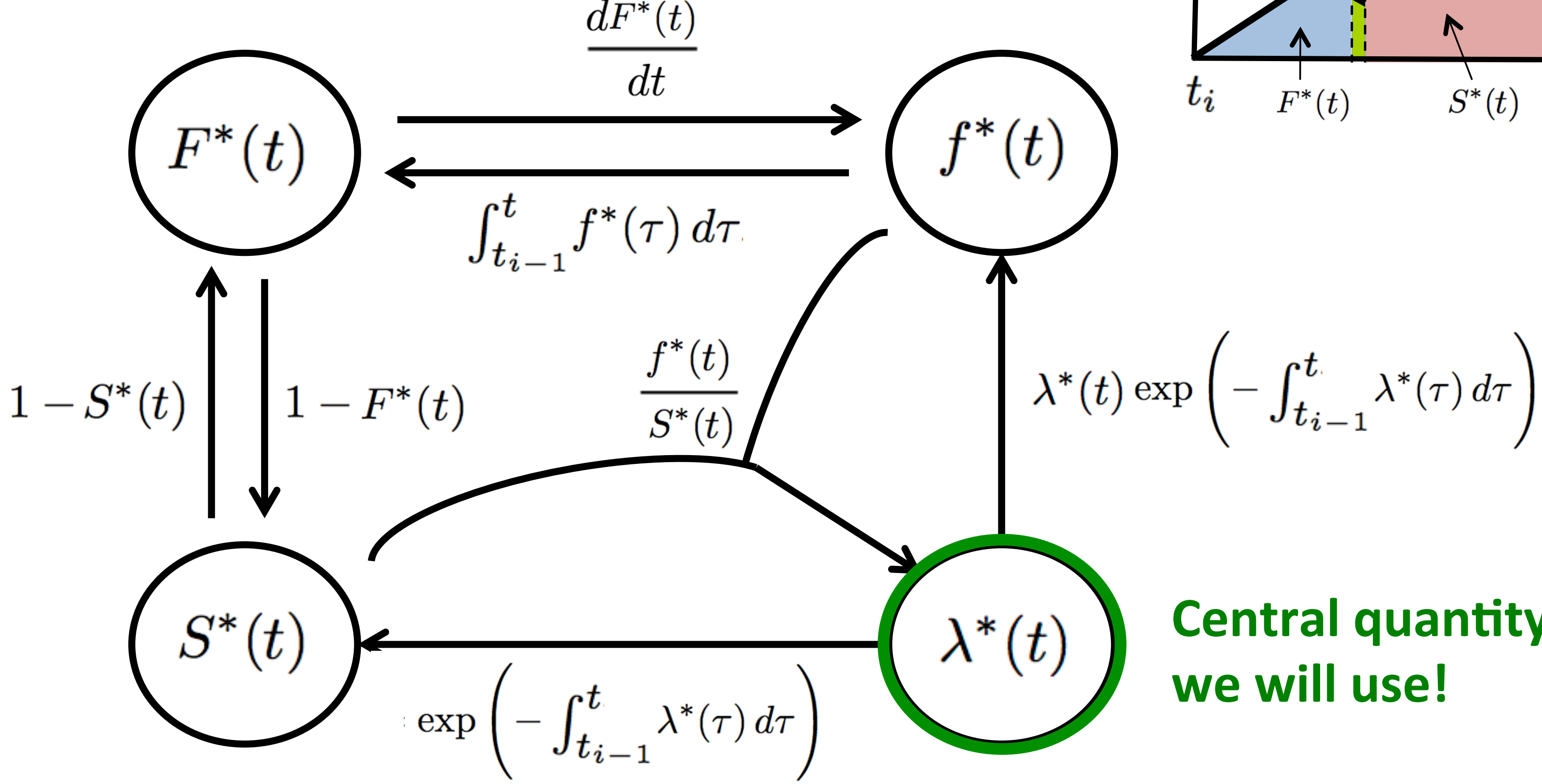
Sum of random processes

$$\lambda^*(t) = \lambda_1^*(t) + \lambda_2^*(t)$$

$$\lambda^*(t) \neq \lambda_1^*(t) * \lambda_2^*(t)$$

Relation between f*, F*, S*, λ*



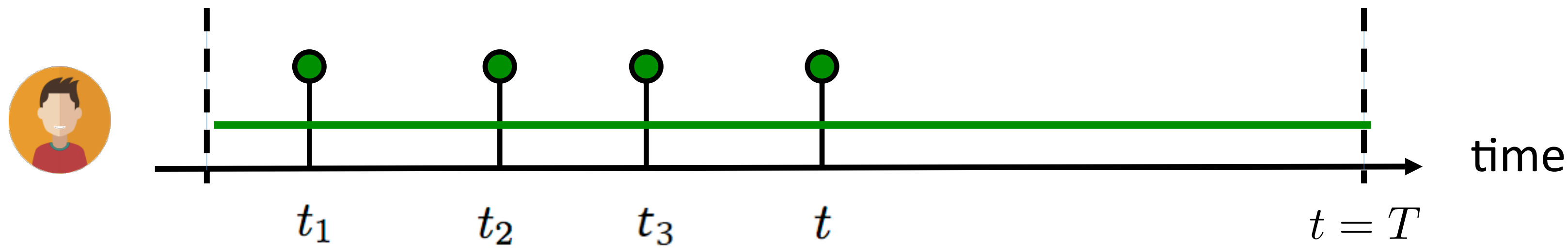


Central quantity we will use!

Representation:
Temporal Point Processes

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

Poisson process



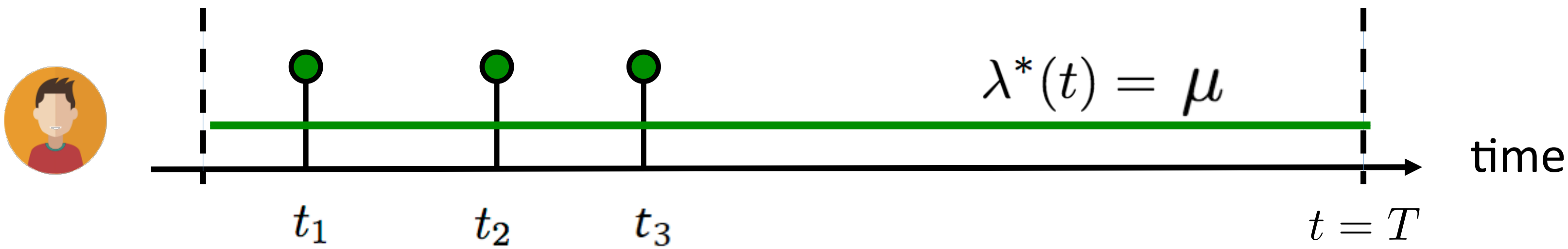
Intensity of a Poisson process

$$\lambda^*(t) = \mu$$

Observations:

1. **Intensity independent of history**
2. **Uniformly random occurrence**
3. **Time interval follows exponential distribution**

Fitting a Poisson from (historical) timeline

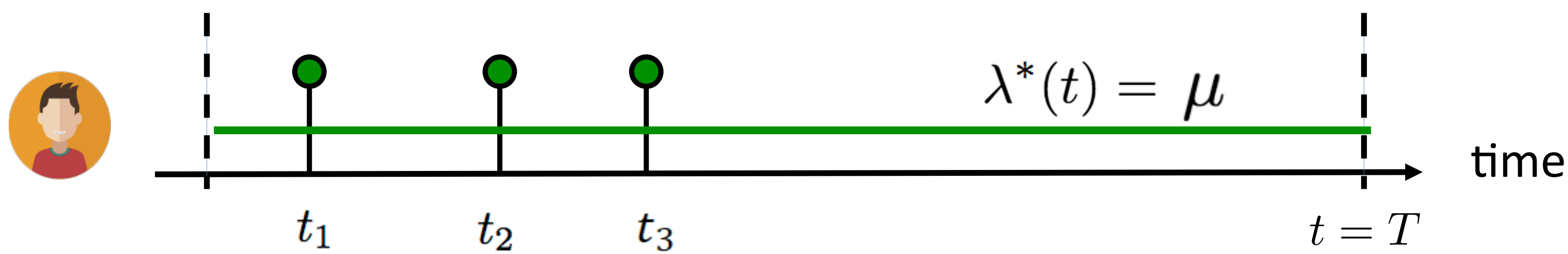


$$\underset{\overset{\uparrow}{\mu}}{\lambda^*(t_1)}\;\underset{\overset{\uparrow}{\mu}}{\lambda^*(t_2)}\;\underset{\overset{\uparrow}{\mu}}{\lambda^*(t_3)}\;\underbrace{\exp\left(-\int_0^T \lambda^*(\tau)\,d\tau\right)}_{\exp(-\mu T)}$$

Maximum likelihood

$$\mu^* = \underset{\mu}{\operatorname{argmax}}\; 3\log\mu - \mu T = \frac{3}{T}$$

Sampling from a Poisson process



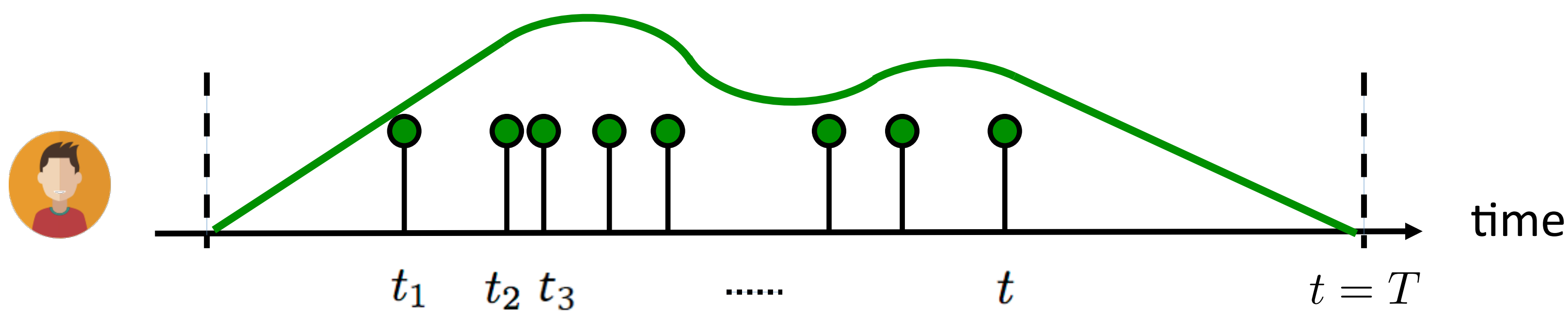
We would like to sample: $t \sim \mu \exp(-\mu(t - t_3)) + t_3$

We sample using inversion sampling:

$$F_t(t) = 1 - \exp(-\mu(t - t_3)) \Rightarrow t \sim \underbrace{-\frac{1}{\mu} \log(1 - \overset{Uniform(0,1)}{\overset{\downarrow}{u}}) + t_3}_{F_t^{-1}(u)}$$

$$\mathbb{P}(F_t^{-1}(u) \leq t) = \mathbb{P}(u \leq F_t(t)) \underset{F_u(u)=u}{\overset{}{=}} F_t(t)$$

Inhomogeneous Poisson process



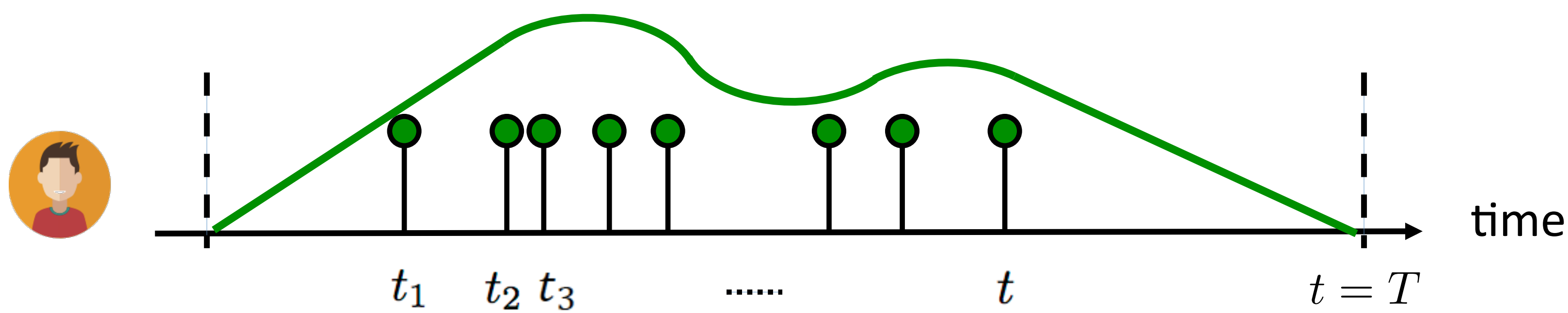
Intensity of an inhomogeneous Poisson process

$$\lambda^*(t) = g(t) \geqslant 0$$

Observations:

1. **Intensity independent of history**

Fitting an inhomogeneous Poisson



$$\lambda^*(t_1)\lambda^*(t_2)\,\lambda^*(t_3)\cdots\cdots\lambda^*(t_n)\;\underbrace{\exp\left(-\int_0^T \lambda^*(\tau)\,d\tau\right)}_{\exp\left(-\int_0^T g(\tau)\,d\tau\right)}$$

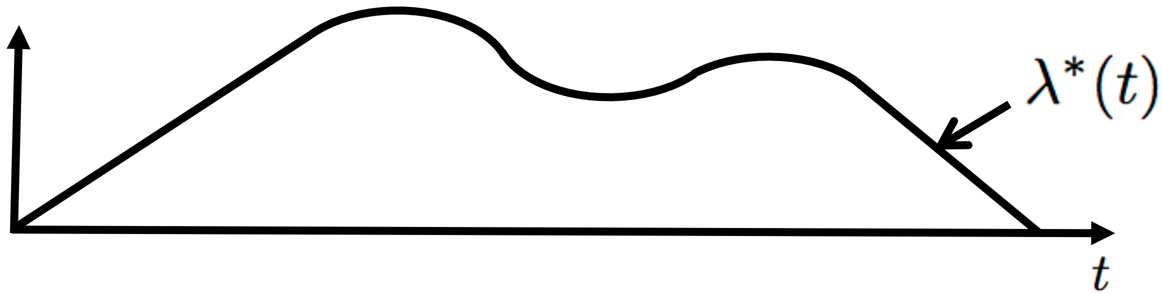
$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ g(t_1) & g(t_2) & g(t_3) & g(t_n) \end{array}$$

Maximum likelihood

$$\underset{g(t)}{\text{maximize}} \quad \sum_{i=1}^{n} \log g(t_i) - \int_0^T g(\tau)\,d\tau$$

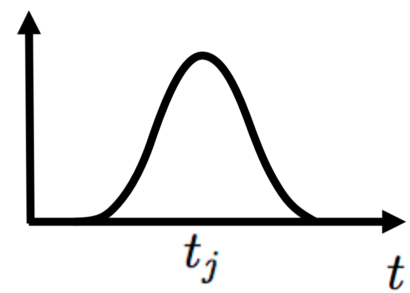
Design $g(t)$ such that max. likelihood is convex (and use CVX)

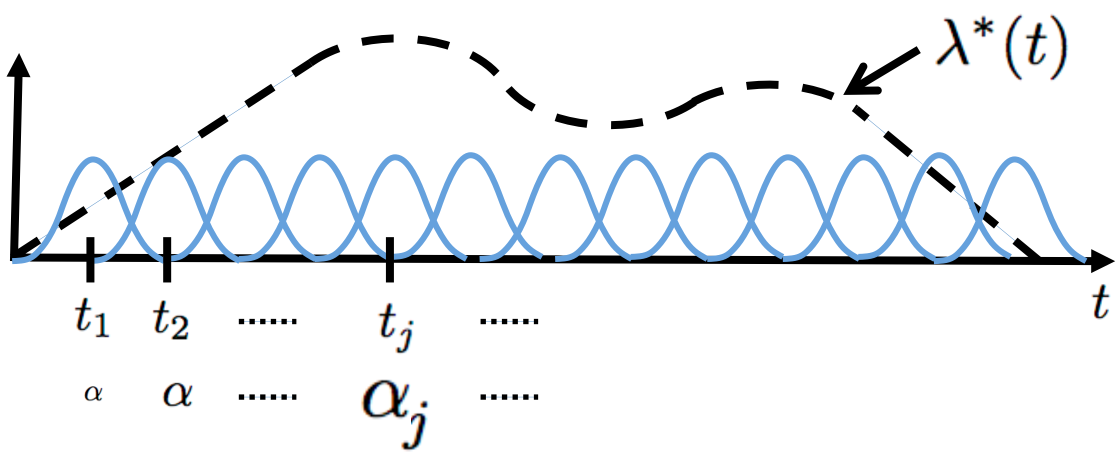
Nonparametric inhomogeneous Poisson process



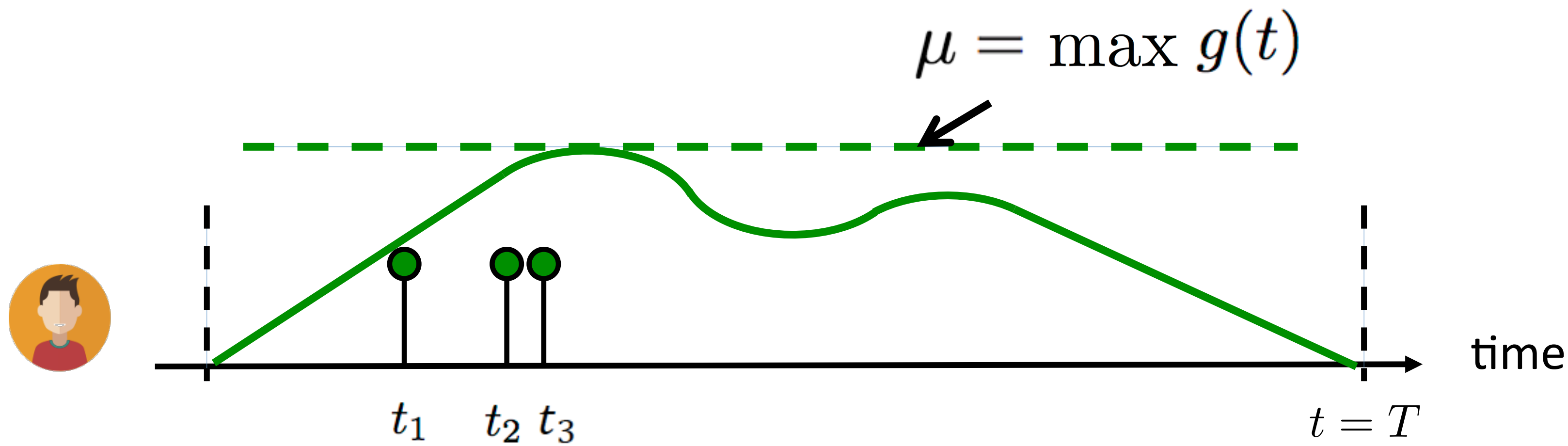
Positive combination of (Gaussian) RFB kernels:

$$\lambda^*(t) = \sum_j \alpha_j \overbrace{k(t - t_j)}$$





Sampling from an inhomogeneous Poisson



Thinning procedure (similar to rejection sampling):

1. **Sample** t **from Poisson process with intensity** μ

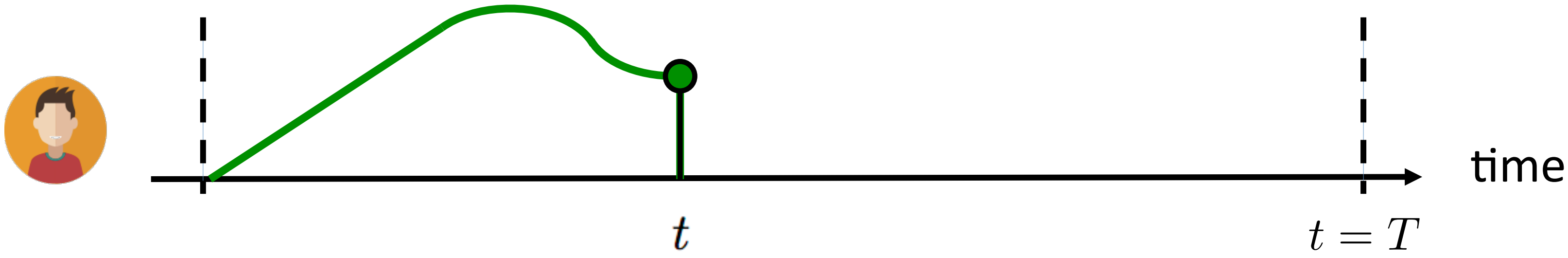
$$t \sim -\frac{1}{\mu} \log(1 - \overset{\overset{Uniform(0,1)}{\downarrow}}{u}) + t_3$$

Inversion sampling

2. **Generate** $u_2 \sim Uniform(0,1)$
3. **Keep the sample if** $u_2 \leq g(t) / \mu$

Keep sample with prob. $g(t)/\mu$

Terminating (or survival) process



Intensity of a terminating (or survival) process

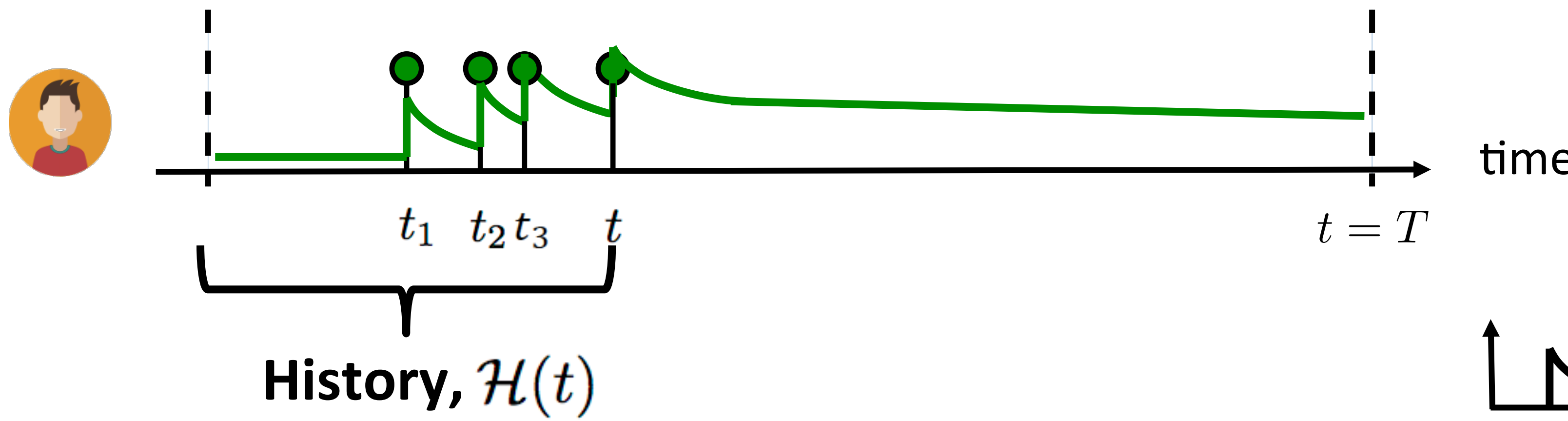
$$\lambda^*(t) = g^*(t)(1 - N(t)) \geqslant 0$$

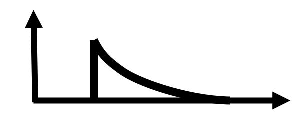
Observations:

1. **Limited number of occurrences**

Try sampling and fitting!

Self-exciting (or Hawkes) process





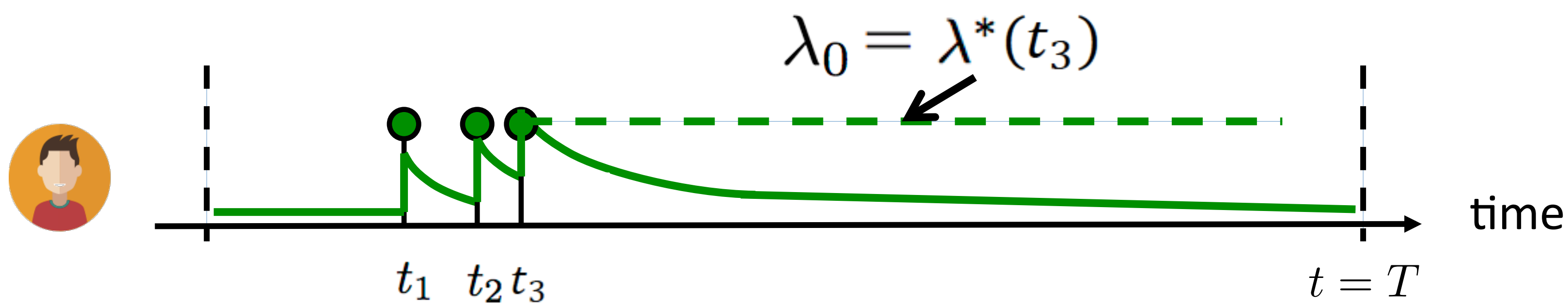
Intensity of self-exciting (or Hawkes) process:

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \overbrace{\kappa_\omega(t - t_i)}^{\text{Triggering kernel}}$$
$$= \mu + \alpha \kappa_\omega(t) \star dN(t)$$

Observations:

1. **Clustered (or bursty) occurrence of events**
2. **Intensity is stochastic and history dependent**

Fitting a Hawkes process from a recorded timeline



$$\lambda^*(t_1)\lambda^*(t_2)\,\lambda^*(t_3)\cdots\cdots\lambda^*(t_n)\;\exp\left(-\int_0^T \lambda^*(\tau)\,d\tau\right)$$

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \kappa_\omega(t - t_i)$$

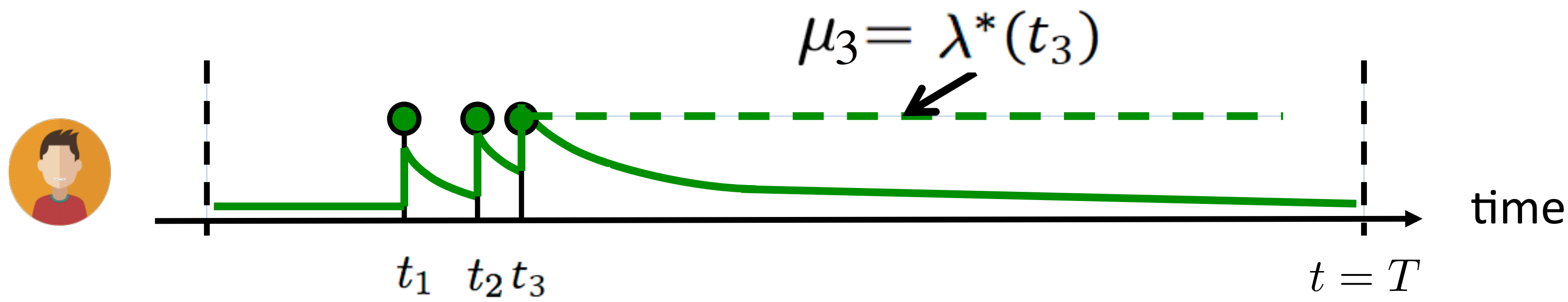
Maximum likelihood

$$\underset{\mu,\alpha}{\text{maximize}} \quad \sum_{i=1}^{n} \log \lambda^*(t_i) - \int_0^T \lambda^*(\tau)\,d\tau$$

The max. likelihood is jointly convex in μ and α

(use CVX!)

Sampling from a Hawkes process



Thinning procedure (similar to rejection sampling):

1. **Sample t from Poisson process with intensity μ_3**

$$t \sim -\frac{1}{\mu_3} \log(1 - \underset{\substack{\uparrow \\ Uniform(0,1)}}{u}) + t_3$$

Inversion sampling

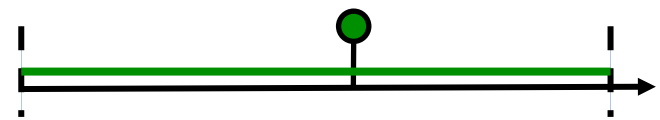
2. **Generate** $u_2 \sim Uniform(0,1)$
3. **Keep the sample if** $u_2 \leq g(t) / \mu_3$

Keep sample with prob. $g(t)/\mu_3$

Summary

Building blocks to represent **different dynamic processes:**

Poisson processes:

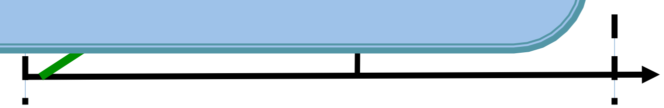


$$\lambda^*(t) = \lambda$$

Inho

We know **how to fit** them
and **how to sample** from them

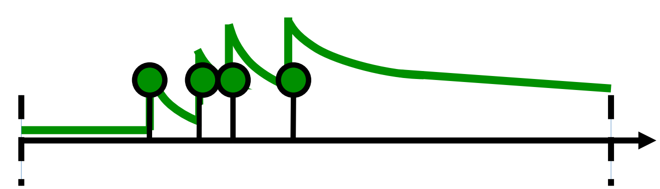
Term



$$\lambda^*(t) = g^*(t)(1 - N(t))$$

Self-exciting point processes:

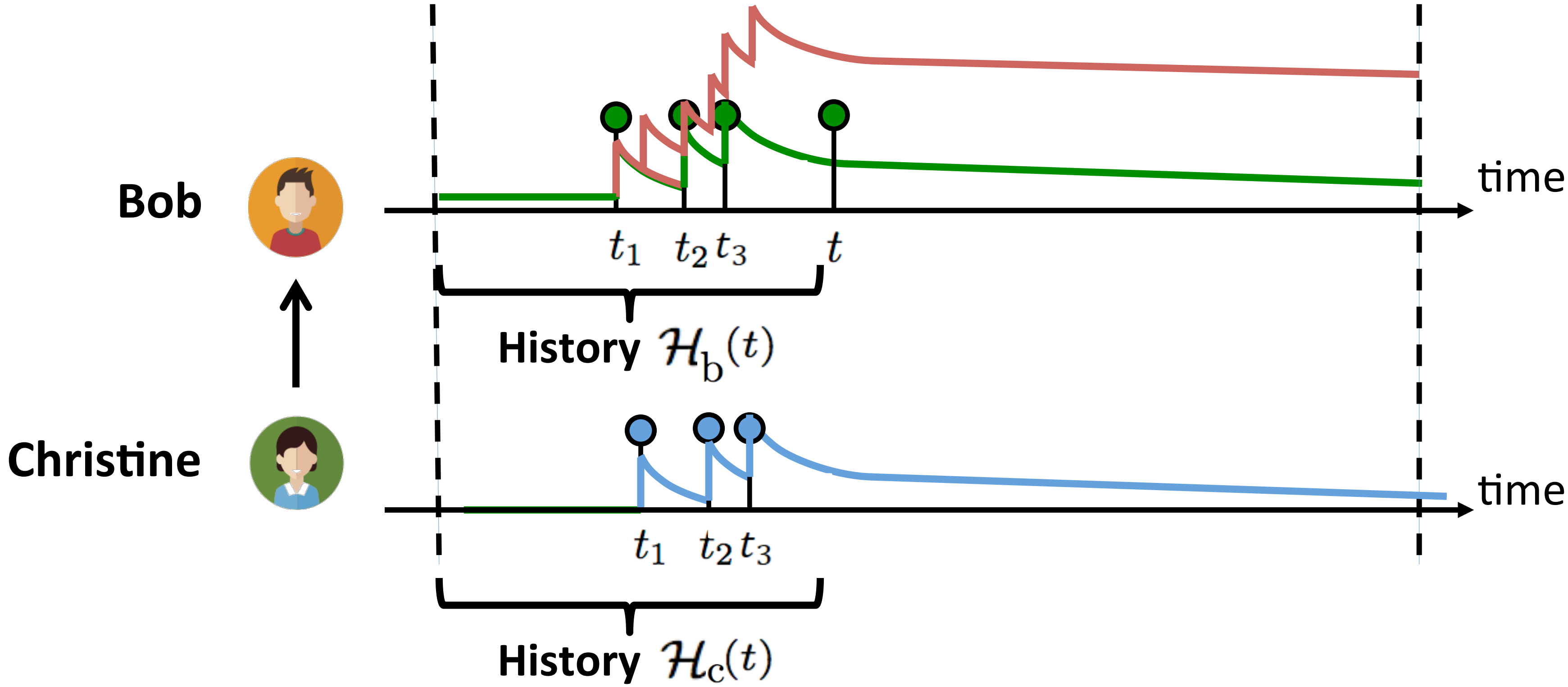
$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \kappa_\omega(t - t_i)$$



Representation:
Temporal Point Processes

1. Intensity function
2. Basic building blocks
3. **Superposition**
4. Marks and SDEs with jumps

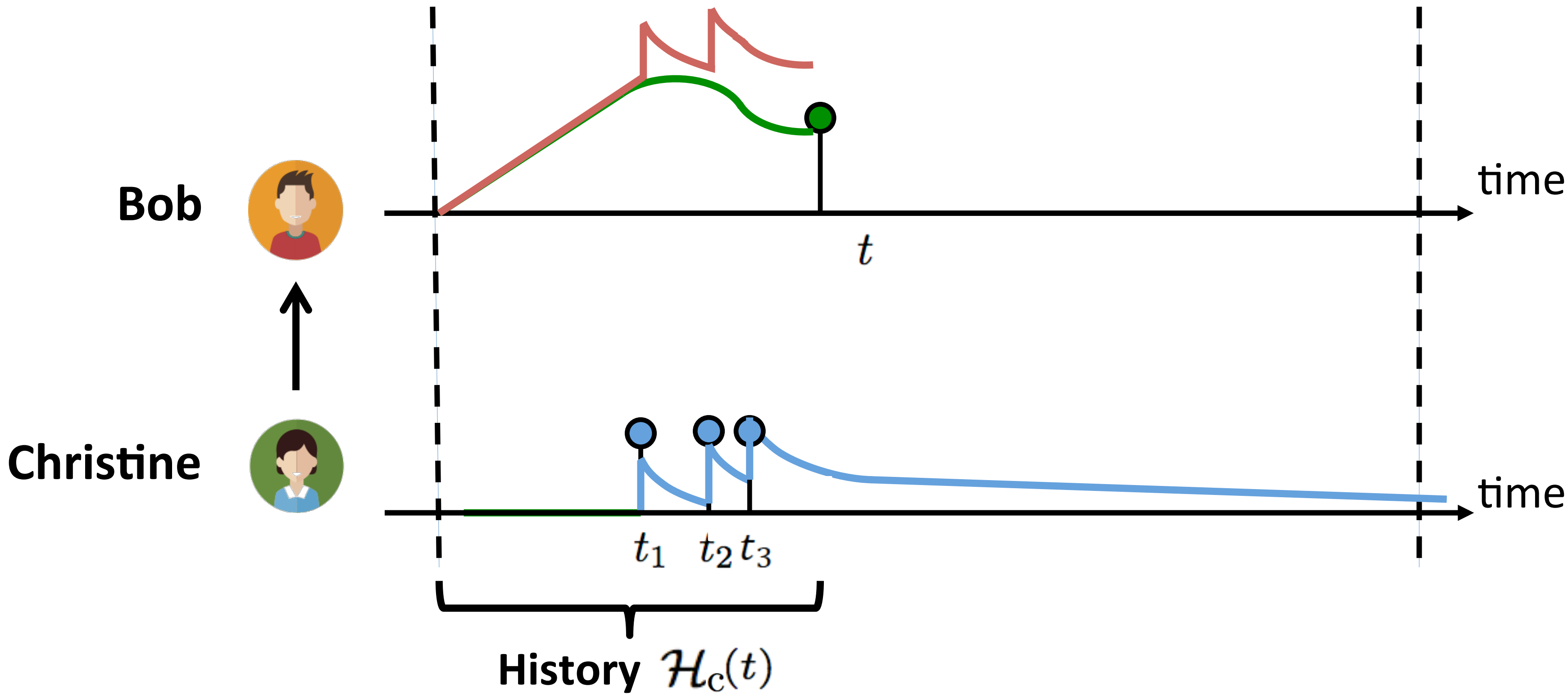
Mutually exciting process



Clustered occurrence affected by neighbors

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}_{\mathrm{b}}(t)} \kappa_\omega(t - t_i) + \beta \sum_{t_i \in \mathcal{H}_{\mathrm{c}}(t)} \kappa_\omega(t - t_i)$$

Mutually exciting terminating process



Clustered occurrence affected by neighbors

$$\lambda^*(t) = (1 - N(t))\left(g(t) + \beta \sum_{t_i \in \mathcal{H}_c(t)} \kappa_\omega(t - t_i)\right)$$

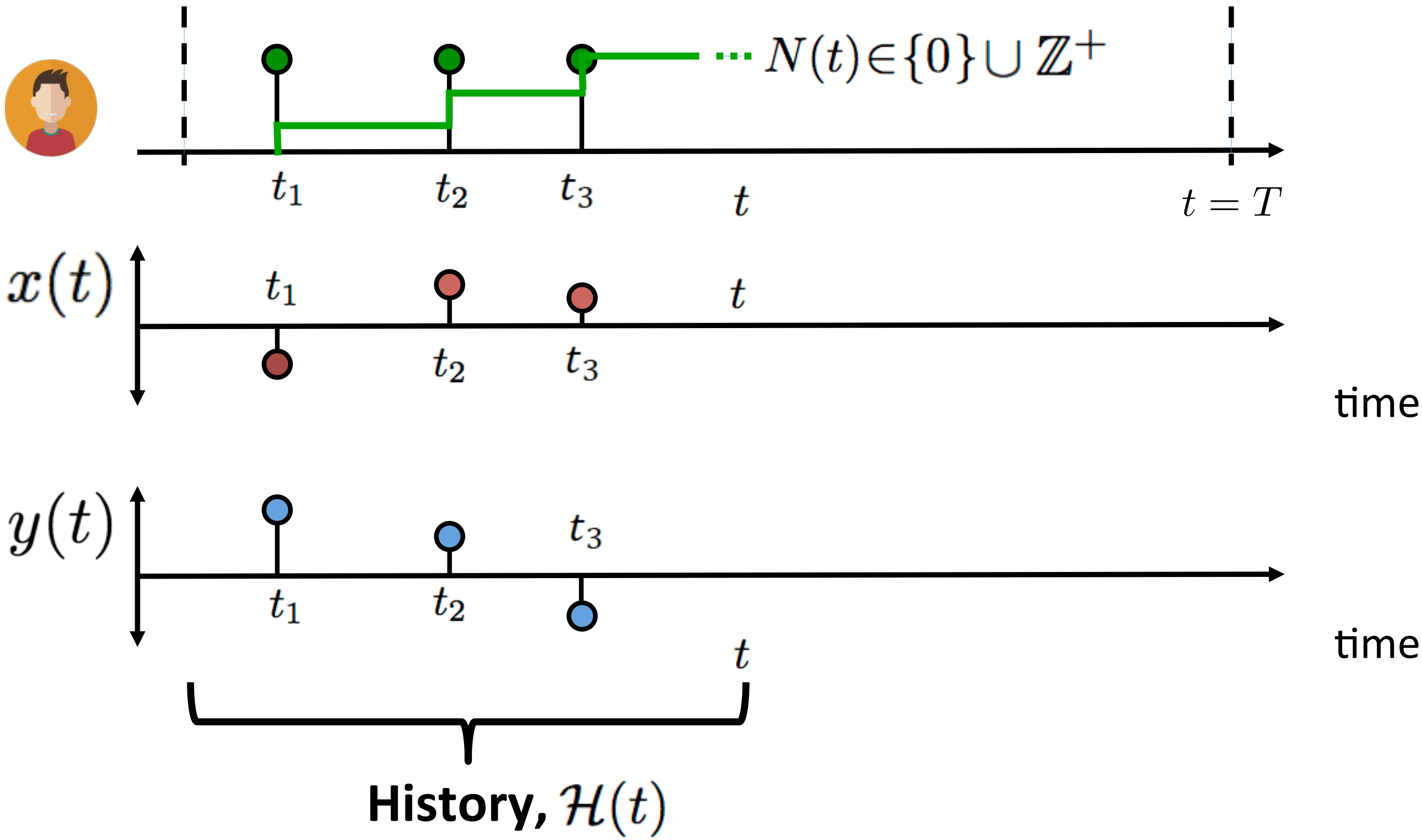
Representation:
Temporal Point Processes

1. Intensity function
2. Basic building blocks
3. Superposition
4. **Marks and SDEs with jumps**

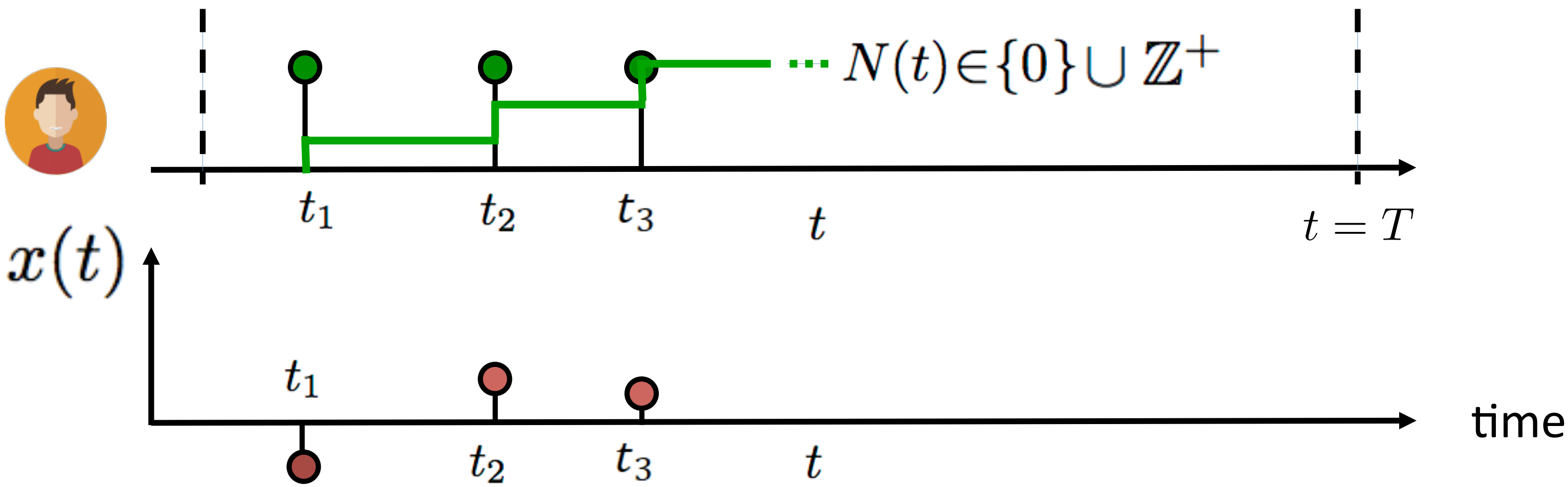
Marked temporal point processes

Marked temporal point process:

A random process whose realization consists of **discrete *marked* events localized in time**



Independent identically distributed marks



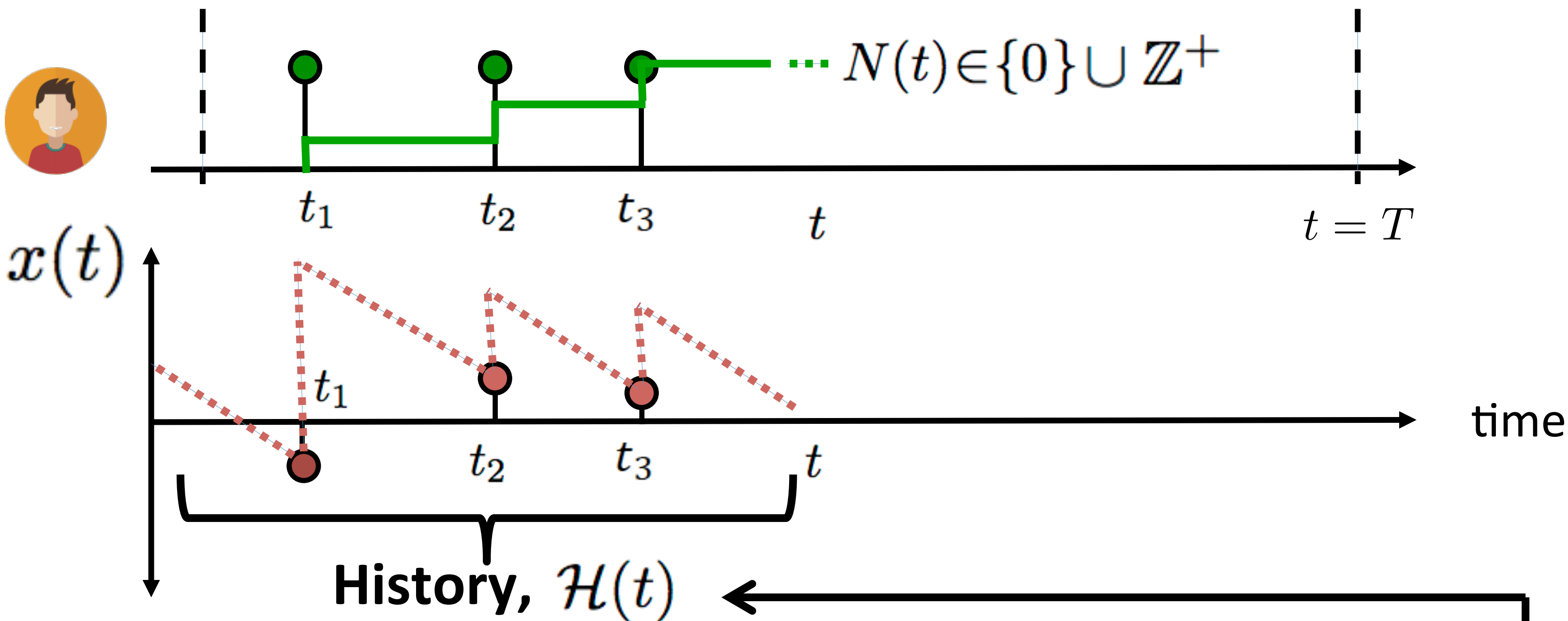
Distribution for the marks:

$$x^*(t_i) \sim p(x)$$

Observations:

1. **Marks independent of the temporal dynamics**
2. **Independent identically distributed (I.I.D.)**

Dependent marks: SDEs with jumps



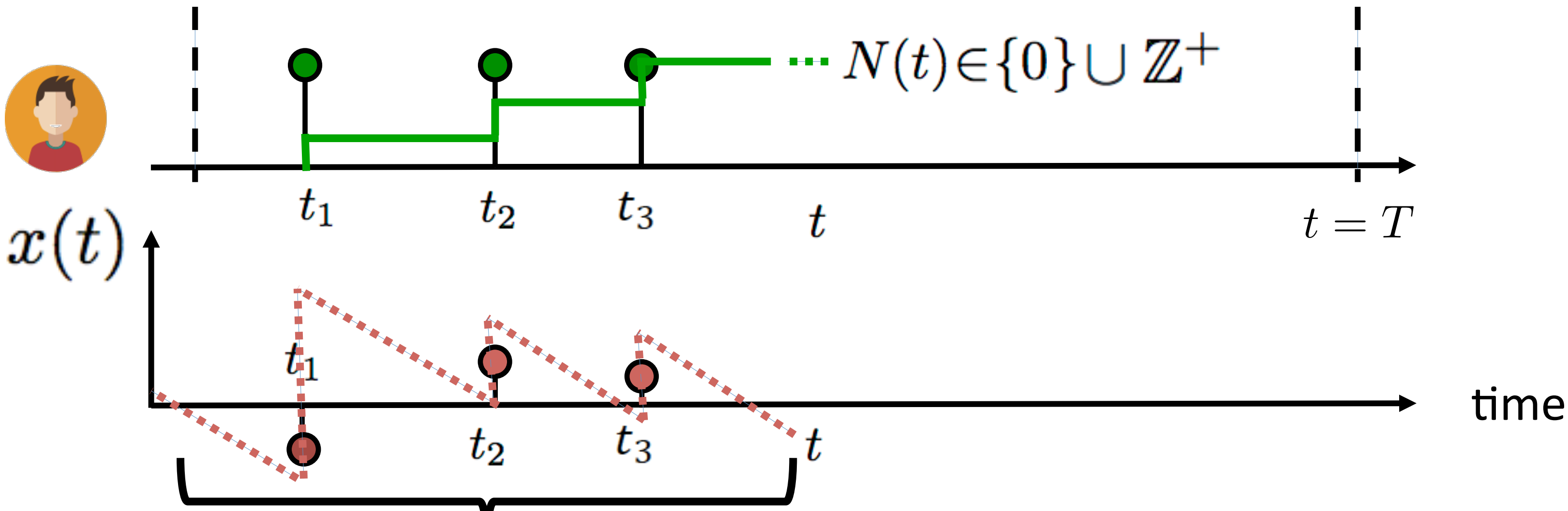
Marks given by stochastic differential equation with jumps:

$$x(t+dt) - x(t) = dx(t) = \underbrace{f(x(t),t)dt}_{\text{Drift}} + \underbrace{h(x(t),t)dN(t)}_{\text{Event influence}}$$

Observations:

1. **Marks dependent of the temporal dynamics**
2. **Defined for all values of t**

Dependent marks: distribution + SDE with jumps



History, $\mathcal{H}(t)$

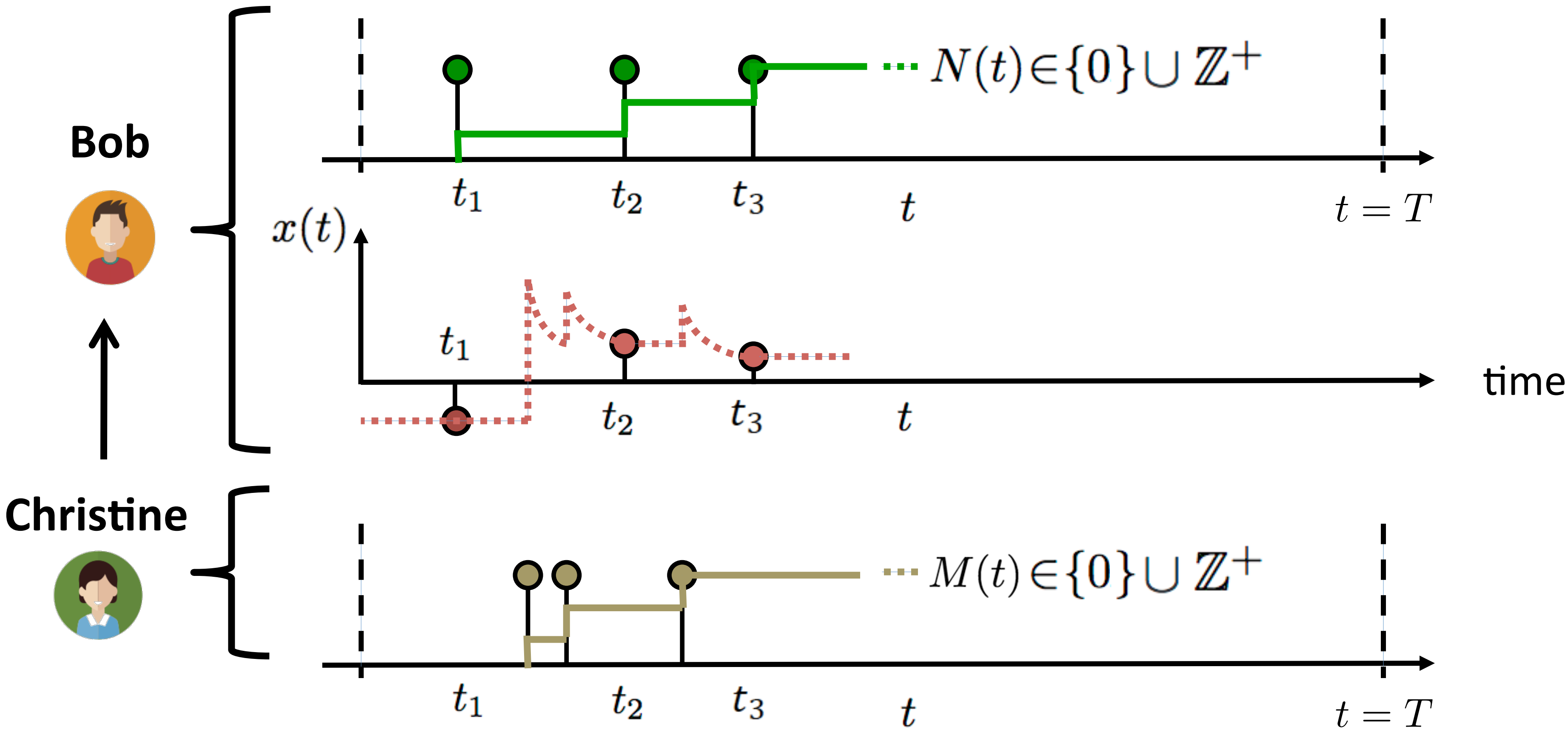
Distribution for the marks:

$$x^*(t_i) \sim p(x^* \mid x(t)) \Rightarrow dx(t) = \underbrace{f(x(t),t)dt}_{\textbf{Drift}} + \underbrace{h(x(t),t)dN(t)}_{\textbf{Event influence}}$$

Observations:

1. **Marks dependent on the temporal dynamics**
2. **Distribution represents additional source of uncertainty**

Mutually exciting + marks



Marks affected by neighbors

$$dx(t) = \underbrace{f(x(t), t)dt}_{\textbf{Drift}} + \underbrace{g(x(t), t)dM(t)}_{\textbf{Neighbor influence}}$$

Representation: Temporal Point processes

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

Applications: Models

1. Information propagation
2. Information reliability
3. Knowledge acquisition

Next

Applications: Control

1. Activity shaping
2. When-to-post

Slides/references: learning.mpi-sws.org/ijcai-2017-tutorial