Statistical Machine Learning

Week 03 (Version 2.0)

Temporal Point Process

Hamid R. Rabiee

Spring 2023

Recall: Binomial Distribution and its relation to Poisson Distribution

Bionomial Distribution: $X \sim B(n, p)$
probability of exactly k success in n trials:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$B(n, p) \quad \xrightarrow[np \text{ remains constant}]{n \to \infty} \quad Poisson(np)$$
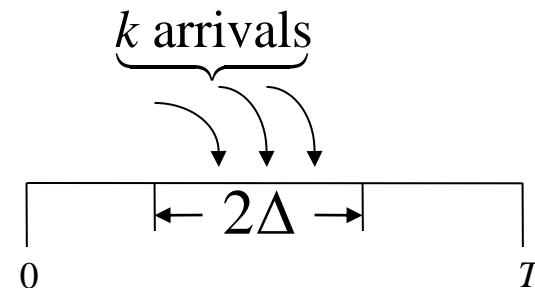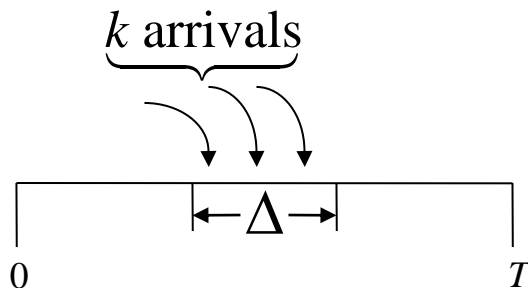
Poisson Processes

- Recall: Binomial and Poisson distributions:
  Both distributions can be used to model the number of occurrences of some event.
- Recall: Poisson arrivals are the limiting behavior of Binomial random variables. (Refer to Poisson approximation of Binomial random variables in your text book):

$$P\left\{\begin{array}{l} \text{"}k \text{ arrivals occur in an} \\ \text{interval of duration } \Delta\text{"} \end{array}\right\} = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \cdots$$

Where:
$$\lambda = np = \mu T \cdot \frac{\Delta}{T} = \mu\Delta$$

Poisson Processes

It follows that:

$$P\left\{\begin{array}{l} "k \text{ arrivals occur in an} \\ \text{interval of duration } 2\Delta" \end{array}\right\} = e^{-2\lambda}\,\frac{(2\lambda)^k}{k!}, \qquad k = 0,\ 1,\ 2,\cdots,$$

since in that case:

$$np_1 = \mu T \cdot \frac{2\Delta}{T} = 2\mu\Delta = 2\lambda.$$

Poisson Processes

- Poisson arrivals over an interval form a Poisson random variable whose parameter depends on the duration of that interval.
- Moreover because of the Bernoulli nature of the underlying basic random arrivals, events over nonoverlapping intervals are independent.
- We shall use these two key observations to define a Poisson process formally.

Poisson Processes

Definition: $X(t) = n(0, t)$ represents a Poisson process if:

(i) the number of arrivals $n(t_1, t_2)$ in an interval (t_1, t_2) of length $t = t_2 - t_1$ is a Poisson random variable with parameter λt. Thus:

$$P\{n(t_1, t_2) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad k = 0, 1, 2, \dots, t = t_2 - t_1$$

And:

Poisson Processes

(ii) If the intervals (t_1, t_2) and (t_3, t_4) are nonoverlapping, then the random variables $n(t_1, t_2)$ and $n(t_3, t_4)$ are independent.

Since $n(0, t) \sim P(\lambda t)$ we have:

$$E[X(t)] = E[n(0,t)] = \lambda t$$

And:

$$E[X^2(t)] = E[n^2(0,t)] = \lambda t + \lambda^2 t^2$$

Poisson Processes

To determine the autocorrelation function $R_{xx}(t_1, t_2)$ let $t_2 > t_1$ then from (ii) above $n(0, t_1)$ and $n(t_1, t_2)$ are independent Poisson random variables with parameters λt_1 and $\lambda(t_2 - t_1)$ respectively.

Thus:

$$E[n(0, t_1)n(t_1, t_2)] = E[n(0, t_1)]E[n(t_1, t_2)] = \lambda^2 t_1(t_2 - t_1)$$

But:
$$n(t_1, t_2) = n(0, t_2) - n(0, t_1) = X(t_2) - X(t_1)$$

And:
$$E[X(t_1)\{X(t_2) - X(t_1)\}] = R_{xx}(t_1, t_2) - E[X^2(t_1)]$$

We obtain:

$$R_{xx}(t_1, t_2) = \lambda^2 t_1(t_2 - t_1) + E[X^2(t_1)] = \lambda t_1 + \lambda^2 t_1 t_2$$
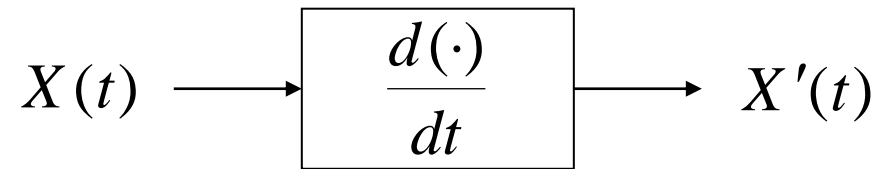$$t_2 \geq t_1$$

Similarly:
$$R_{xx}(t_1, t_2) = \lambda t_2 + \lambda^2 t_1 t_2$$

Thus:
$$R_{xx}(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda \min(t_1, t_2)$$

Example:

$$X(t) \longrightarrow \boxed{\dfrac{d(\cdot)}{dt}} \longrightarrow X'(t)$$

(Derivative as a LTI system)

Then:

$$\mu_{x'}(t) = \frac{d\mu_x(t)}{dt} = \frac{d\lambda t}{dt} = \lambda, \quad a\ constant$$

And:

$$R_{xx'}(t_1, t_2) = \frac{\partial R_{xx}(t_1,\ t_2)}{\partial t_2} = \begin{cases} \lambda^2 t_1 & t_1 \le t_2 \\ \lambda^2 t_1 + \lambda & t_1 > t_2 \end{cases}$$

$$= \lambda^2 t_1 + \lambda\ U(t_1 - t_2)$$

And:

$$R_{x'x'}(t_1,\ t_2) = \frac{\partial R_{xx'}(t_1, t_2)}{\partial t_1} = \lambda^2 + \lambda\ \delta(t_1 -\ t_2).$$

Poisson Processes

Notice that:

- The Poisson process $X(t)$ *does not* represent a wide sense stationary process.

- Although $X(t)$ *does not* represent a wide sense stationary process, its derivative $X'(t)$ *does* represent a wide sense stationary process.

Poisson Processes

Since $X'(t)$ is a wide sense stationary process;
nonstationary inputs to linear systems *can* lead to
wide sense stationary outputs, an interesting observation.

- **Sum of Poisson Processes:**

  If $X_1(t)$ and $X_2(t)$ represent two independent Poisson processes,
  then their sum $X_1(t) + X_2(t)$ is also a Poisson process with
  parameter $(\lambda_1 + \lambda_2)t$. (Follows from the definition
  of the Poisson process in (i) and (ii)).

Poisson Processes

Random selection of Poisson Points:

Let $t_1, t_2, \cdots, t_i, \cdots$ represent random arrival points associated with a Poisson process $X(t)$ with parameter λt, and associated with each arrival point, define an independent Bernoulli random variable N_i, where:
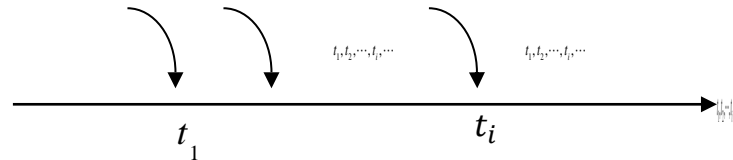
$$P(N_i = 1) = p, \qquad P(N_i = 0) = q = 1 - p.$$

Poisson Processes

Define the processes:

$$Y(t) = \sum_{i=1}^{X(t)} N_i \quad ; \quad Z(t) = \sum_{i=1}^{X(t)} (1 - N_i) = X(t) - Y(t)$$

We claim that both $Y(t)$ and $Z(t)$ are independent Poisson processes with parameters $\lambda p t$ and $\lambda q t$, respectively, where $q = 1 - p$. When X(t) is a Poisson process with parameter λt.

Poisson Processes

Proof:

$$Y(t) = \sum_{n=k}^{\infty} P\{Y(t) = k \mid X(t) = n\} P\{X(t) = n)\}.$$

But given $X(t) = n$, we have $Y(t) = \sum_{i=1}^{n} N_i \sim B(n, p)$ so that:

$$P\{Y(t) = k \mid X(t) = n\} = \binom{n}{k} p^k q^{n-k}, \qquad 0 \le k \le n,$$

And:

$$P\{X(t) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

$$P\{Y(t)=k\} = e^{-\lambda t} \sum_{n=k}^{\infty} \tfrac{n!}{(n-k)!k!} \, p^k \, q^{n-k} \, \tfrac{(\lambda t)^n}{n!} = \frac{p^k e^{-\lambda t}}{k!} \, (\lambda t)^k \underbrace{\sum_{n=k}^{\infty} \tfrac{(q\lambda t)^{n-k}}{(n-k)!}}_{e^{q\lambda t}}$$

$$= (\lambda pt)^k \, \frac{e^{-(1-q)\lambda t}}{k!} = e^{-\lambda pt} \, \frac{(\lambda pt)^k}{k!}, \qquad k=0,\,1,\,2,\,\cdots$$

$$\sim \quad P(\lambda pt).$$

More generally:

$$P\{Y(t)=k, Z(t)=m\} = P\{Y(t)=k, X(t)-Y(t)=m\}$$

$$= P\{Y(t)=k, X(t)=k+m\}$$

$$= P\{Y(t)=k \mid X(t)=k+m\} P\{X(t)=k+m\}$$

$$= \binom{k+m}{k} p^k q^m \cdot e^{-\lambda t} \frac{(\lambda t)^{k+m}}{(k+m)!} = \underbrace{e^{-\lambda pt} \frac{(\lambda pt)^n}{k!}}_{P(Y(t)=k)} \underbrace{e^{-\lambda qt} \frac{(\lambda qt)^n}{m!}}_{P(Z(t)=m)}$$

$$= P\{Y(t)=k\} P\{Z(t)=m\},$$

Notice that $Y(t)$ and $Z(t)$ are generated as a result of random Bernoulli selections from the original Poisson process $X(t)$, where each arrival gets tossed over to either $Y(t)$ with probability p or to $Z(t)$ with probability q. Each such sub-arrival stream is also a Poisson process. Thus random selection of Poisson points preserve the Poisson nature of the resulting processes. However, deterministic selection from a Poisson process destroys the Poisson property for the resulting processes.
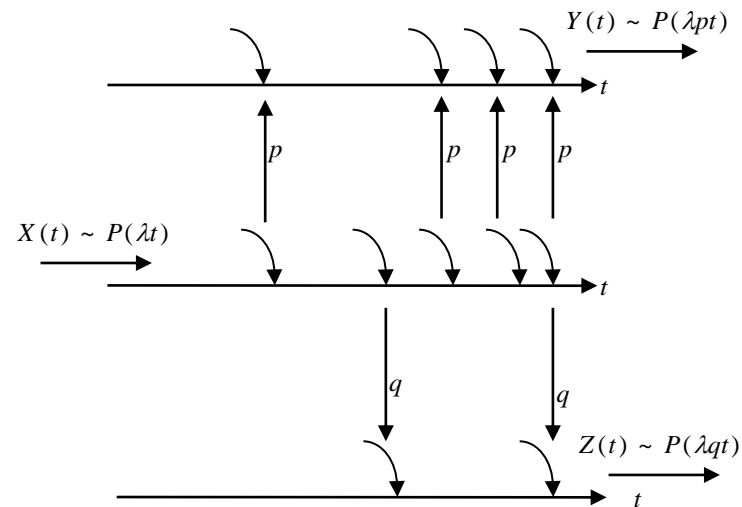
Inter-arrival Distribution for Poisson Processes

Let τ_1 denote the time interval (delay) to the first arrival from *any* fixed point t_0. To determine the probability distribution of the random variable τ_1, we argue as follows: Observe that the event "$\tau_1 > t$" is the same as "$n(t_0, t_0+t) = 0$", or the complement event "$\tau_1 \leq t$" is the same as the event "$n(t_0, t_0+t) > 0$" .
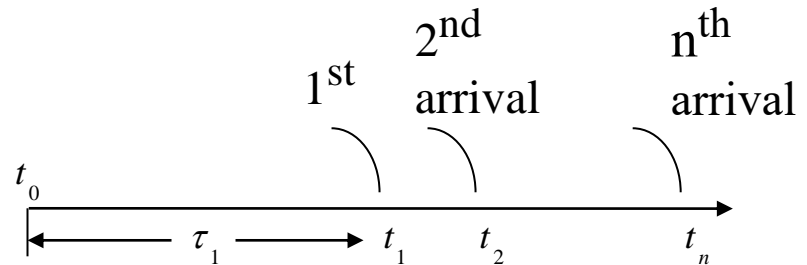
Inter-arrival Distribution for Poisson Processes

Hence the distribution function of τ_1 is given by:

$$F_{\tau_1}(t) \triangleq P\{\tau_1 \leq t\} = P\{X(t) > 0\} = P\{n(t_0, t_0 + t) > 0\}$$

$$= 1 - P\{n(t_0, t_0 + t) = 0\} = 1 - e^{-\lambda t}$$

Its derivative gives the probability density function for τ_1 to be:

$$f_{\tau_1}(t) = \frac{dF_{\tau_1}(t)}{dt} = \lambda e^{-\lambda t}, \qquad t \geq 0$$

i.e. τ_1 is an exponential random variable with parameter λ so that: $E(\tau_1) = 1/\lambda$.

Inter-arrival Distribution for Poisson Processes

Similarly, let t_n represent the nth random arrival point for a Poisson process. Then:

$$\triangle F_{t_n}(t) = P\{t_n \le t\} = P\{X(t) \ge n\}$$

$$= 1 - P\{X(t) < n\} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

and hence:

$$f_{t_n}(x) = \frac{dF_{t_n}(x)}{dx} = -\sum_{k=1}^{n-1} \frac{\lambda(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x} + \sum_{k=0}^{n-1} \frac{\lambda(\lambda x)^k}{k!} e^{-\lambda x}$$

$$= \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, \quad x \ge 0$$

Inter-arrival Distribution for Poisson Processes

which represents a Gamma density function. i.e., the waiting time to the nth Poisson arrival has a Gamma distribution. Moreover:

$$t_n = \sum_{i=1}^{n} \tau_i$$

where τ_i is the random inter-arrival duration between the $(i-1)^{th}$ and i^{th} events. Notice that τ_i s are independent, identically distributed random variables. Hence using their characteristic functions, it follows that all inter-arrival durations of a Poisson process are independent exponential random variables with common parameter λ. i.e.,

$$f_{\tau_i}(t) = \lambda \, e^{-\lambda t}, \qquad t \geq 0.$$

Inter-arrival Distribution for Poisson Processes

Alternatively, we have τ_1 is an exponential random variable.
By repeating that argument after shifting t_0 to the
new point t_1, we conclude that τ_2 is an exponential random variable.
Thus the sequence $\tau_1, \tau_2, \cdots, \tau_n, \cdots$ are independent exponential random variables with common p.d.f.

Thus if we systematically tag every m^{th} outcome of a Poisson process $X(t)$ with parameter λt to generate a new process $e(t)$, then the inter-arrival time between any two events of $e(t)$ is a Gamma random variable.

Inter-arrival Distribution for Poisson Processes

Notice that:

$$E[e(t)] = m / \lambda, \text{ and if } \lambda = m\mu, \text{ then } E[e(t)] = 1/\mu.$$

The inter-arrival time of $e(t)$ in that case represents an Erlang-m random variable, and $e(t)$ is an Erlang-m process.
In summary, if Poisson arrivals are randomly redirected to form new queues, then each such queue generates a new Poisson process.

Poisson Departures between Exponential Inter-arrivals

Let $X(t) \sim P(\lambda t)$ and $Y(t) \sim P(\mu t)$ represent two independent Poisson processes called *arrival* and *departure* processes.
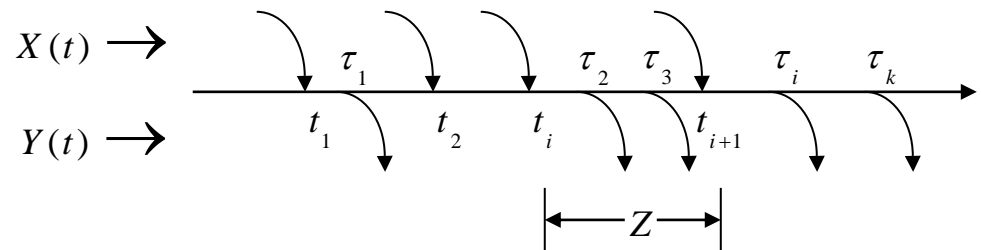
Let Z represent the random interval between *any* two successive arrivals of $X(t)$. Z has an exponential distribution with parameter λ. Let N represent the number of "departures" of $Y(t)$ between *any* two successive arrivals of $X(t)$. Then from the Poisson nature of the departures we have:

$$P\{N = k \mid Z = t\} = e^{-\mu t} \frac{(\mu t)^k}{k!}.$$

Poisson Departures between Exponential Inter-arrivals

$$P\{N = k\} = \int_0^\infty P\{N = k \mid Z = t\} f_z(t)\, dt$$

$$= \int_0^\infty e^{-\mu t}\, \frac{(\mu t)^k}{k!}\, \lambda e^{-\lambda t}\, dt$$

$$= \frac{\lambda}{k!} \int_0^\infty (\mu t)^k\, e^{-(\lambda + \mu)t}\, dt$$

$$= \frac{\lambda}{\lambda + \mu} \left(\frac{\mu}{\lambda + \mu} \right)^k \frac{1}{k!} \underbrace{\int_0^\infty x^k e^{-x}\, dx}_{k!}$$

$$= \left(\frac{\lambda}{\lambda + \mu} \right) \left(\frac{\mu}{\lambda + \mu} \right)^k, \qquad k = 0,\, 1,\, 2,\, \cdots$$

.

Poisson Departures between Exponential Inter-arrivals

The random variable N has a geometric distribution. Thus if customers come in and get out according to two independent Poisson processes at a counter, then the number of arrivals between any two departures has a geometric distribution. Similarly the number of departures between *any* two arrivals also represents another geometric distribution.

Example

Suppose there are 2 Poisson processes with $\lambda_1 = 1, \lambda_2 = 2$.
Find the probability that 2nd arrival of first process occurs before 3rd arrival of the second process.

Solution:
Consider the superposition of these two Poisson processes. It is still a Poisson process with $\lambda = 1 + 2 = 3$. Also each event of the resulting process is from first process with probability $\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{1}{3}$ and otherwise with probability $\frac{2}{3}$. So for the 2nd arrival of first process to occur before 3rd arrival of the second process, we need the first 4 occurrences to cover at least 2 occurrences of the first process:

$$\sum_{k=2}^{4} \binom{4}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{4-k}$$

Example: Coupon Collecting

Suppose a cereal manufacturer inserts a sample of one type of coupon randomly into each cereal box. Suppose there are n such distinct types of coupons. One interesting question is that how many boxes of cereal should one buy on the average in order to collect at least one coupon of each kind?

Example: Coupon Collecting

We shall reformulate the above problem in terms of Poisson processes. Let $X_1(t), X_2(t), \cdots, X_n(t)$ represent *n independent* identically distributed Poisson processes with common parameter λt. Let $t_{i1}, t_{i2}, \cdots$ represent the first, second, … random arrival instants of the process $X_i(t)$, $i = 1, 2, \cdots, n$. They will correspond to the first, second, $\cdots$ appearance of the ith type coupon in the above problem. Let:

$$X(t) \triangleq \sum_{i=1}^{n} X_i(t),$$

so that the sum $X(t)$ is also a Poisson process with parameter μt, where

$$\mu = n\lambda.$$

Example: Coupon Collecting

$1/\lambda$ represents: The average inter-arrival duration between any two arrivals of $X_i(t), i = 1, 2, \cdots, n,$ whereas:

$1/\mu$ represents the average inter-arrival time for the combined sum process $X(t)$.
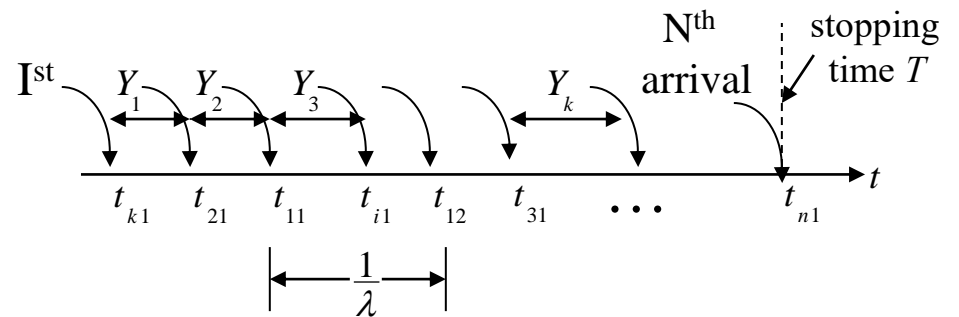
Bulk Arrivals and Compound Poisson Processes

In an ordinary Poisson process $X(t)$, only one event occurs at any arrival instant. Instead suppose a random number of events C_i occur simultaneously as a cluster at every arrival instant of a Poisson process. If $X(t)$ represents the total number of all occurrences in the interval $(0, t)$, then $X(t)$ represents a **compound Poisson process**, or a **bulk arrival process**.

Bulk Arrivals and Compound Poisson Processes

Inventory orders, arrivals at an airport queue, tickets purchased for a show, etc. follow this process (when things happen, they happen in a bulk, or a bunch of items are involved.)
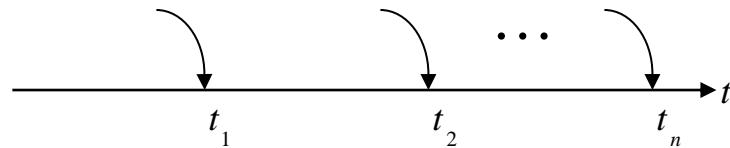
(a) Poisson Process
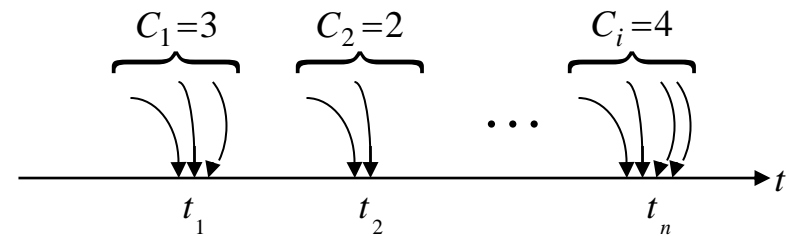
(b) Compound Poisson Process

Let:

$$p_k = P\{C_i = k\}, \quad k = 0, 1, 2, \cdots$$

represent the common probability mass function for the occurrence in any cluster C_i. Then the compound process $X(t)$ satisfies:

$$X(t) = \sum_{i=1}^{N(t)} C_i,$$

where $N(t)$ represents an ordinary Poisson process with parameter λ. Let:

$$P(z) = E\{z^{C_i}\} = \sum_{k=0}^{\infty} p_k z^k$$

represent the moment generating function associated with the cluster Statistics. Then the moment generating function of the compound Poisson process $X(t)$ is given by:

$$\phi_X(z) = \sum_{n=0}^{\infty} z^n P\{X(t) = n\} = E\{z^{X(t)}\}$$

$$= E\{E[z^{X(t)} \mid N(t) = k]\} = E[E\{z^{\sum_{i=1}^{k} C_i} \mid N(t) = k\}]$$

$$= \sum_{k=0}^{\infty} (E\{z^{C_i}\})^k P\{N(t) = k\}$$

$$= \sum_{k=0}^{\infty} P^k(z) e^{-\lambda t} \frac{(\lambda t)^k}{k!} = e^{-\lambda t(1 - P(z))}$$

If we let:

$$P^k(z) \overset{\Delta}{=} \left(\sum_{n=0}^{\infty} p_n z^k \right)^k = \sum_{n=0}^{\infty} p_n^{(k)} z^n$$

where $\{ p_n^{(k)} \}$ represents the k fold convolution of the sequence $\{ p_n \}$ with itself, we obtain:

$$P\{X(t) = n\} = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} p_n^{(k)}$$

The above, represents the probability that there are n arrivals in the interval $(0, t)$ for a compound Poisson process $X(t)$.

We can rewrite $\phi_x(z)$ also as:

$$\phi_X(z) = e^{-\lambda_1 t(1-z)} e^{-\lambda_2 t(1-z^2)} \cdots e^{-\lambda_k t(1-z^k)} \cdots$$

where $\lambda_k = p_k \lambda$, which shows that the compound Poisson process can be expressed as the sum of integer-scaled independent Poisson processes $m_1(t), m_2(t), \cdots$. Thus:

$$X(t) = \sum_{k=1}^{\infty} k\, m_k(t).$$

More generally, every linear combination of independent Poisson processes represents a compound Poisson process.

Point Processes

Many discrete *events* in continuous time
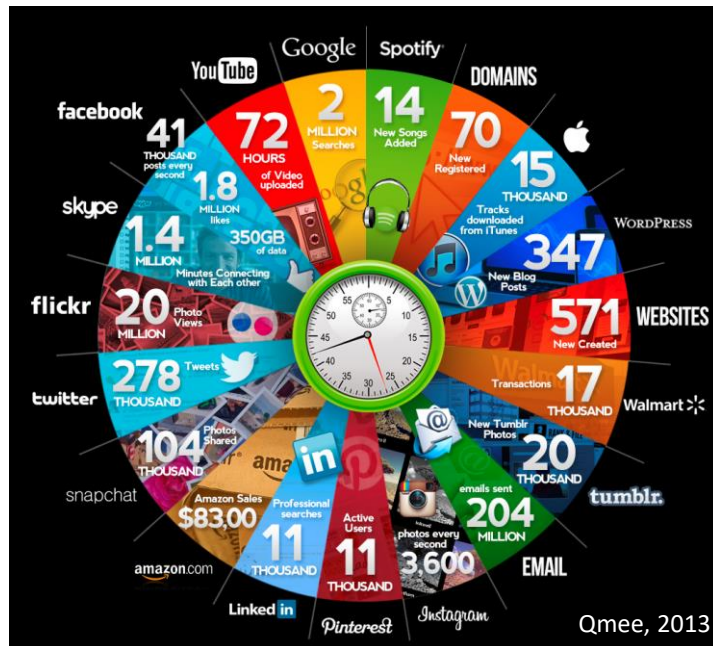
Online actions

Disease dynamics

Financial trading
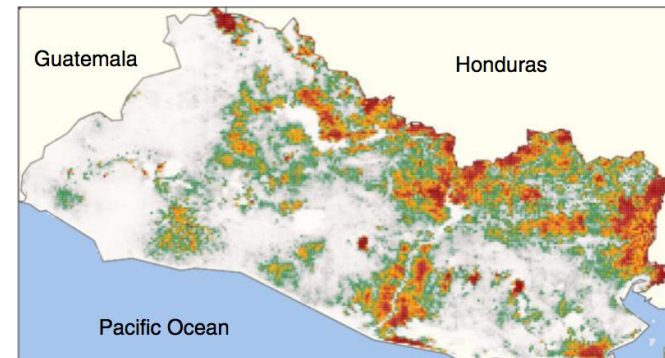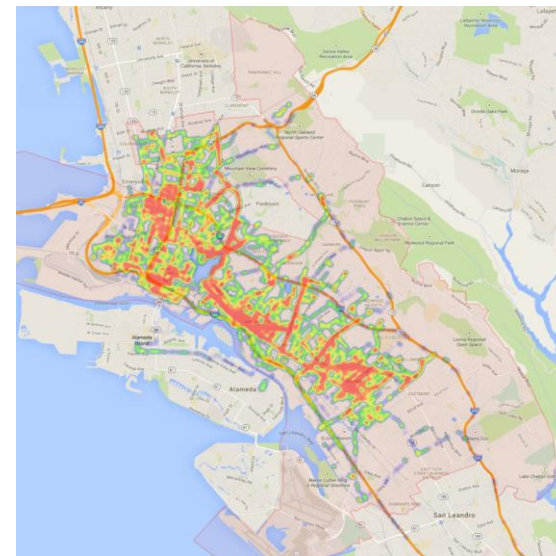
Mobility dynamics

38

Variety of processes behind these events

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

Stock trading

Flu spreading

Article creation in Wikipedia

News spread in Twitter

Reviews and sales in Amazon

Ride-sharing requests

A user's reputation in Quora

FAST → SLOW

...in a wide range of temporal scales.

Example I: Information propagation

S ⟶ D
means
D follows S

Christine — 3.00pm
Bob — 3.25pm
Beth — 3.27pm
Joe
David — 4.15pm

Friggeri et al., 2014
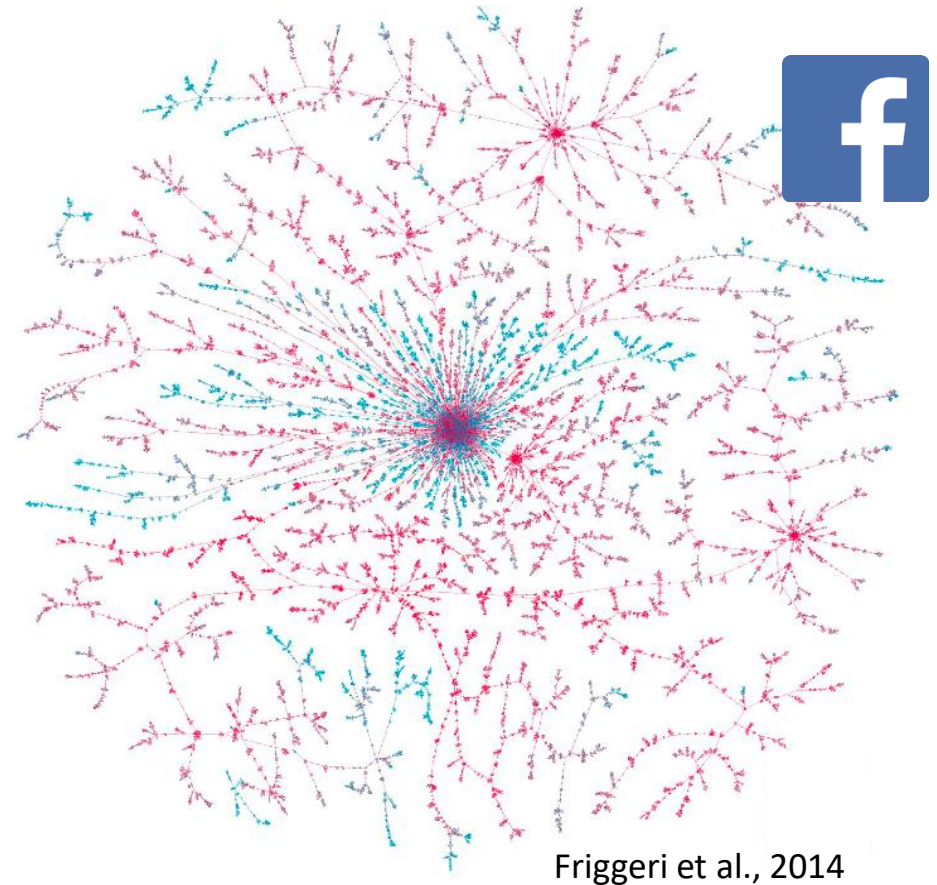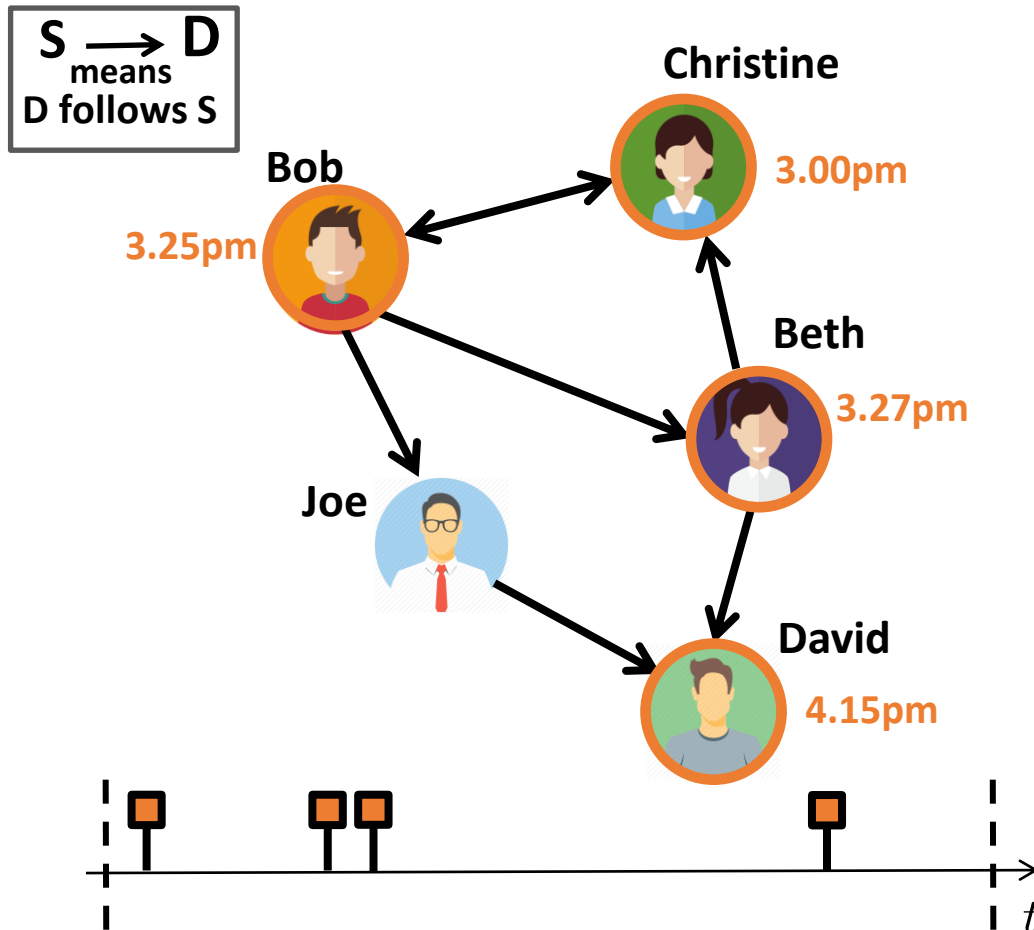
They can have an impact in the off-line world

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

Barack Obama

From Wikipedia, the free encyclopedia

"Barack" and "Obama" redirect here. For his father, see Barack Obama Sr. For other uses of "Barack", see Barack (disambiguation). (disambiguation).

Barack Hussein Obama II (l current President of the Unite continental United States. Bo was president of the Harvard civil rights attorney and taugh representing the 13th District States House of Represental

Barack Obama: Revision history

03:41, 28 November 2016 Ranze (talk | contribs) . . (301,105 bytes) (+18) . . (B

03:32, 28 November 2016 Xin Deui (talk | contribs) . . (301,087 bytes) (-68) . . (

00:57, 28 November 2016 SporkBot (talk | contribs) m . . (301,155 bytes) (-37)

07:03, 27 November 2016 Saiph121 (talk | contribs) . . (301,192 bytes) (+25) . .

03:21, 20 September 2016

is a Kenyan politician

possible vandalism by MLM2016

is an American politician

● **Addition**

■ **Refutation**
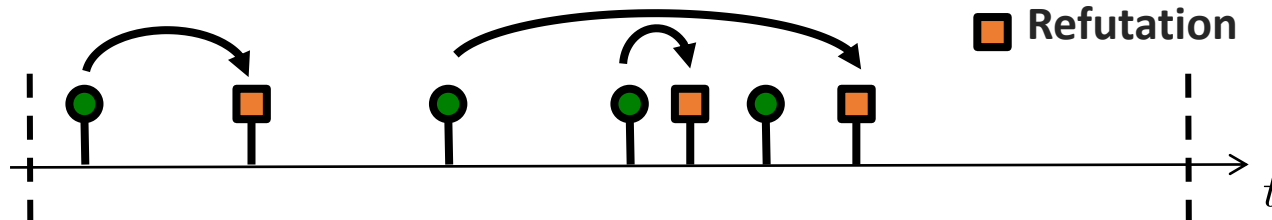
Moving to Australia   Working in Australia   Study abroad in Australia   +4   ✎

What are the pros and cons of living in Australia?

✎ Answer   Request ▾   Follow **109**   Comment   Share **9**   Downvote

I have studied, worked and lived in Australia as an Internal employee, business owner and a citizen.

Upvote | 150

I have experienced this country in all the ways possible, you However, I firmly believe that there are definitely more pros Australia but still I have mentioned below a few . . . . allenges and benefits.

Hope it helps! :)

Possible Challenges

• Language problem for those who don't speak g
• Not having your family and friends around coul
  society is more and more connected and thanks
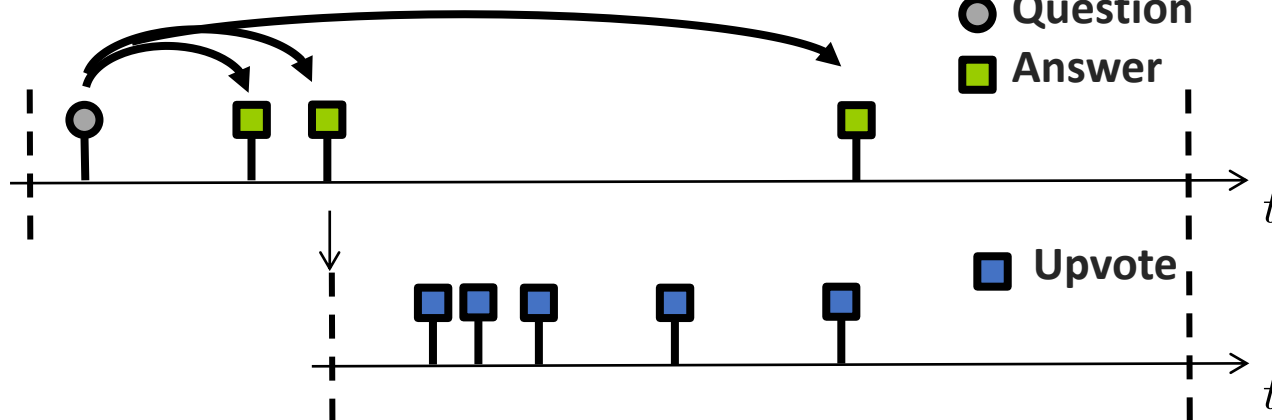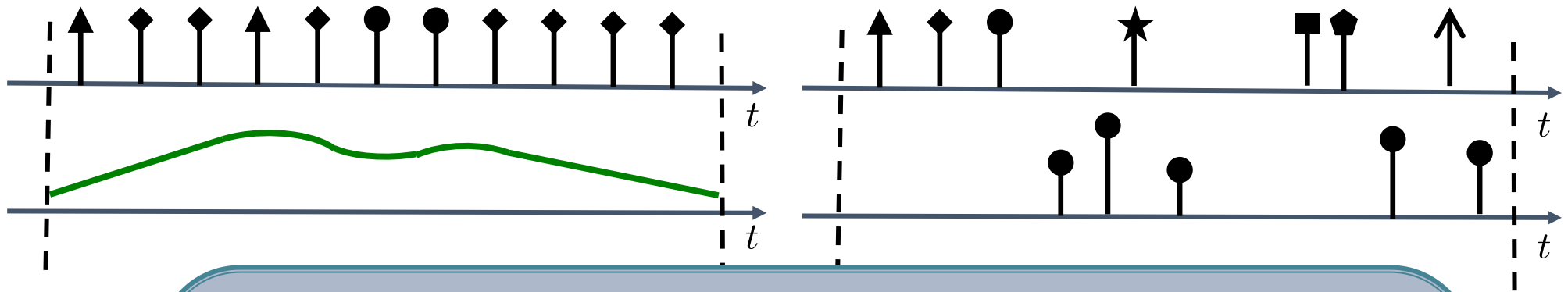 Social Media you can stay in touch a bit easier w

v M Sharma, Lived in Australia as Migrant, Student, Worker, Business Owner & Family Man
Updated Aug 3

● **Question**

■ **Answer**

■ **Upvote**

Aren't these event traces just time series?

The framework of
temporal point processes
provides a *native representation*

Disc

Wh ...och?
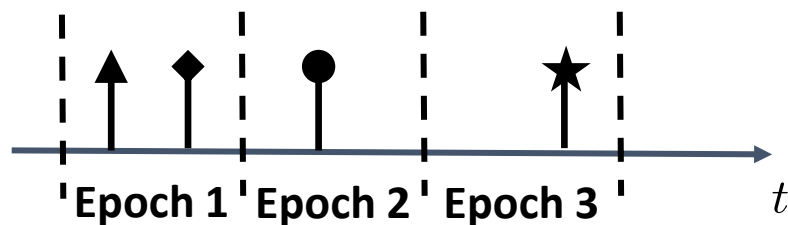
What about time-related queries?

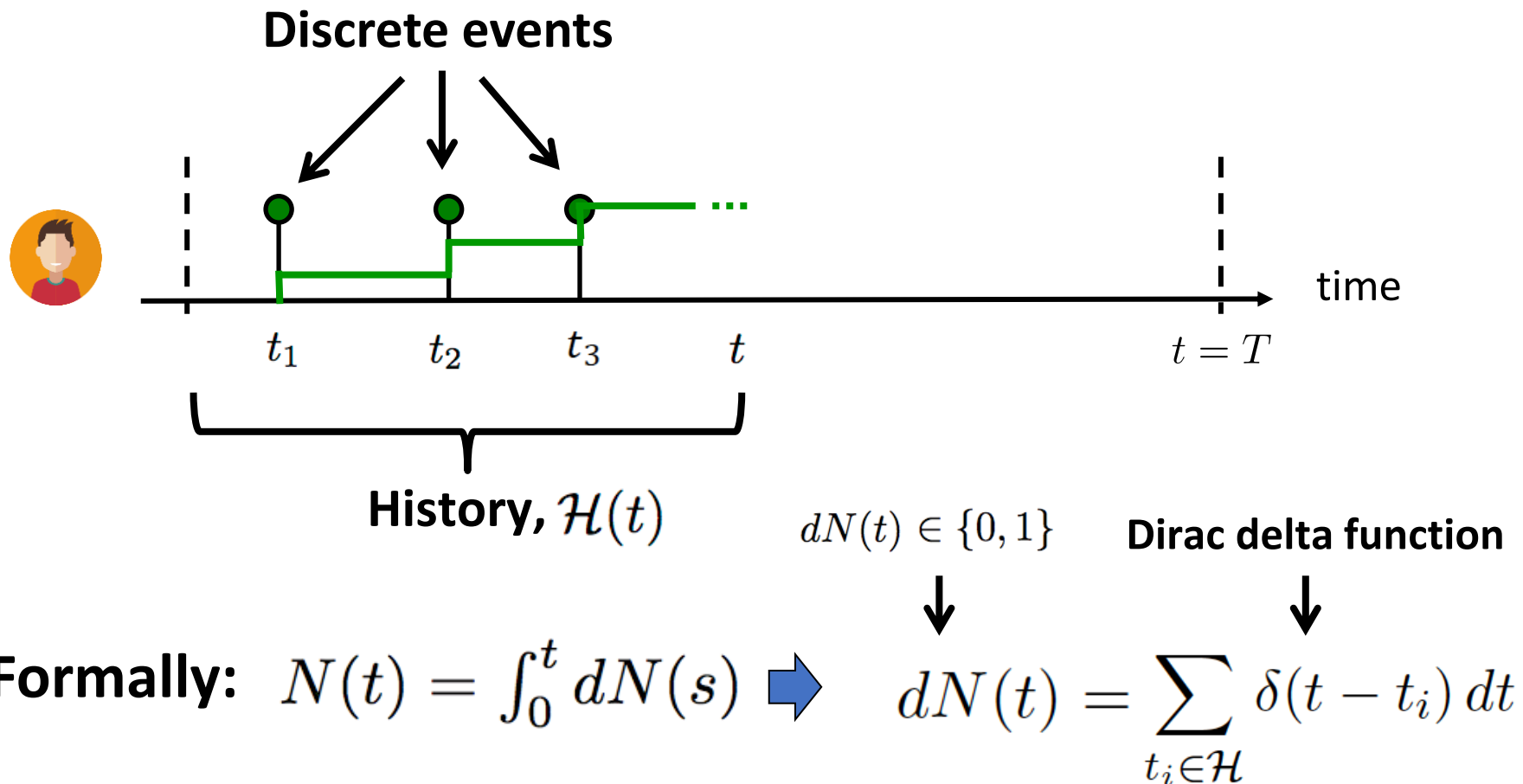'Epoch 1 'Epoch 2' Epoch 3 ' t

Temporal Point Processes (TPPs):
Introduction

1. Intensity function

2. Basic building blocks
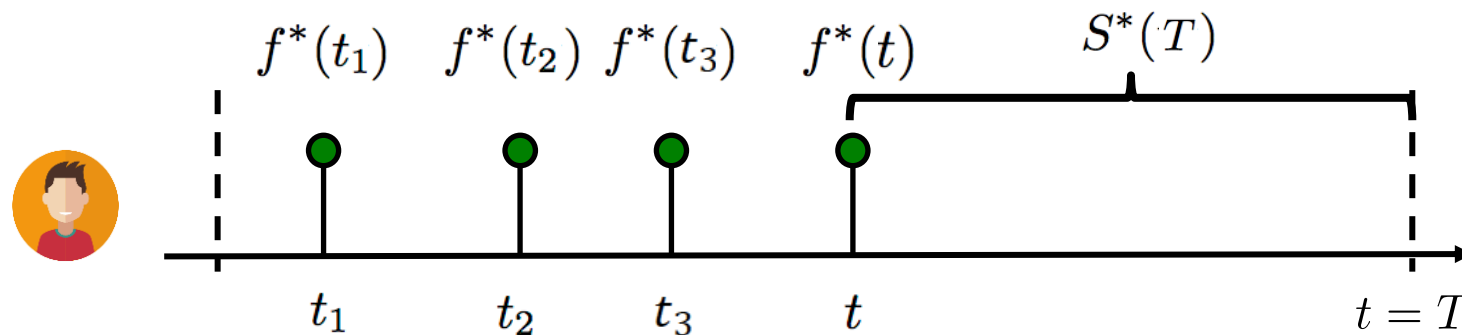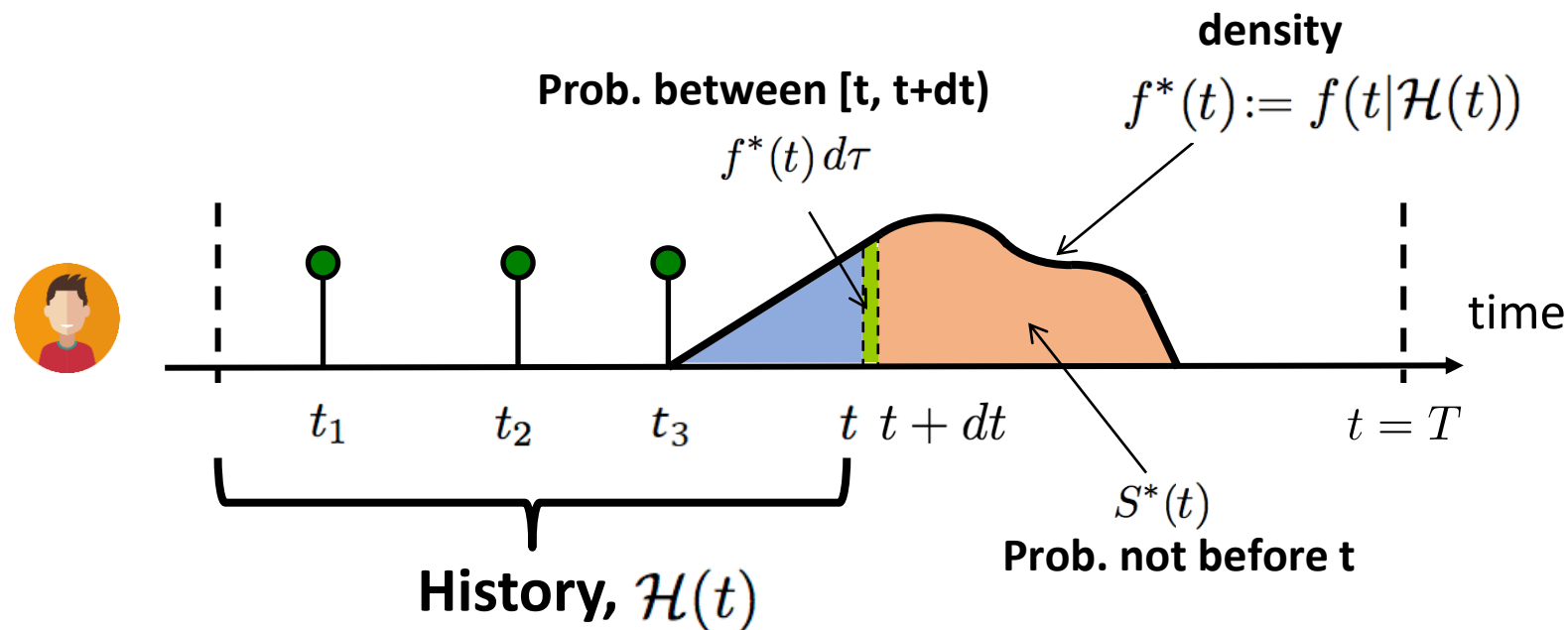
3. Superposition

4. Marks and SDEs with jumps

Temporal point processes

Temporal point process:
A random process whose realization consists of **discrete events localized in time** $\mathcal{H} = \{t_i\}$

Discrete events

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

$dN(t) \in \{0, 1\}$ **Dirac delta function**

Formally: $N(t) = \int_0^t dN(s)$ $\Rightarrow$ $dN(t) = \sum_{t_i \in \mathcal{H}} \delta(t - t_i) \, dt$

Model time as a random variable

Prob. between [t, t+dt)
$$f^*(t)\,d\tau$$

density
$$f^*(t) := f(t|\mathcal{H}(t))$$

time

$t_1 \quad t_2 \quad t_3 \quad t \; t+dt \qquad t = T$

$S^*(t)$
Prob. not before t

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

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

$t_1 \qquad t_2 \qquad t_3 \qquad t \qquad\qquad t = T$

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

45

Problems of density parametrization (I)
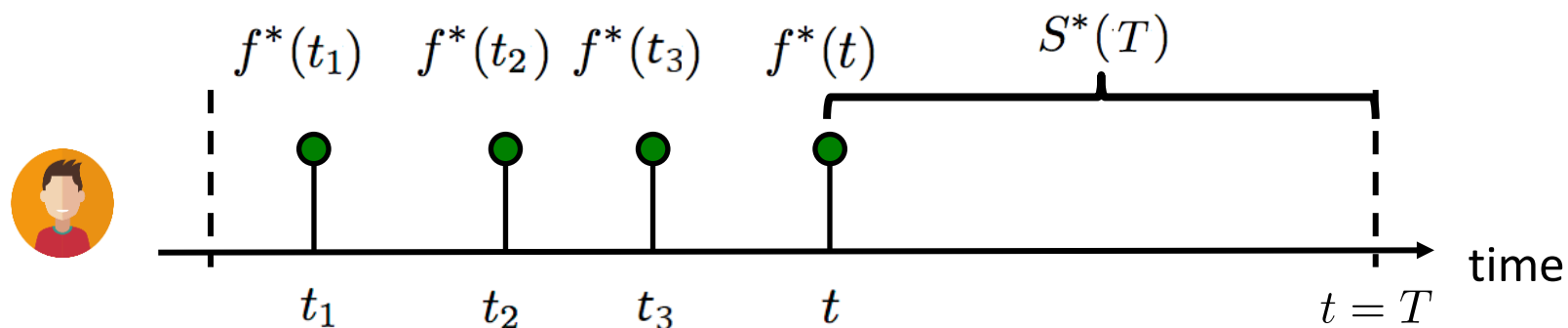
$$f^*(t_1) \quad f^*(t_2) \quad f^*(t_3) \quad f^*(t) \qquad S^*(T)$$

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

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

$$\frac{\exp\langle w, \psi^*(t_2) \rangle}{Z} \qquad \frac{\exp\langle w, \psi^*(t) \rangle}{Z}$$

It is difficult for model design and interpretability:

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

Intensity function

density

$$f^*(t) := f(t|\mathcal{H}(t))$$

Prob. between [t, t+dt)

$$f^*(t)\,d\tau$$

$S^*(t)$
Prob. not before t

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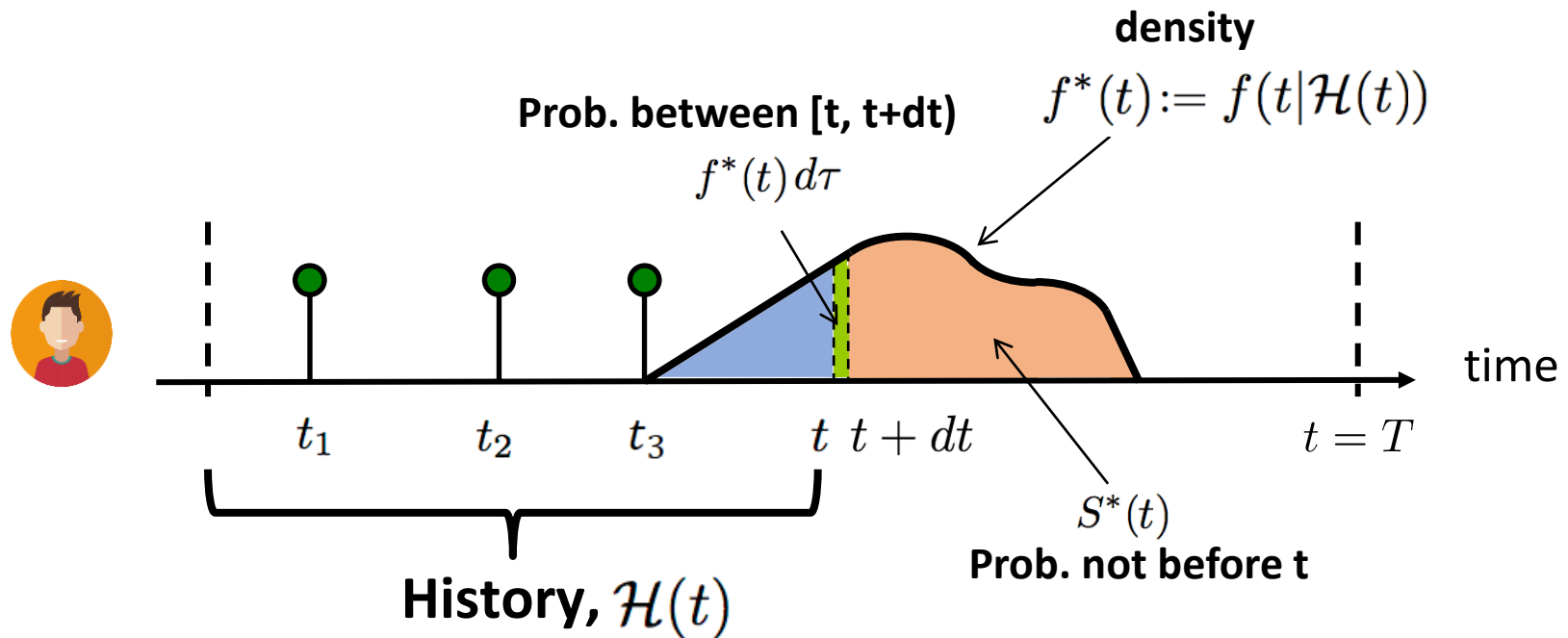
Intensity:

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**

47

Advantages of intensity parametrization (I)
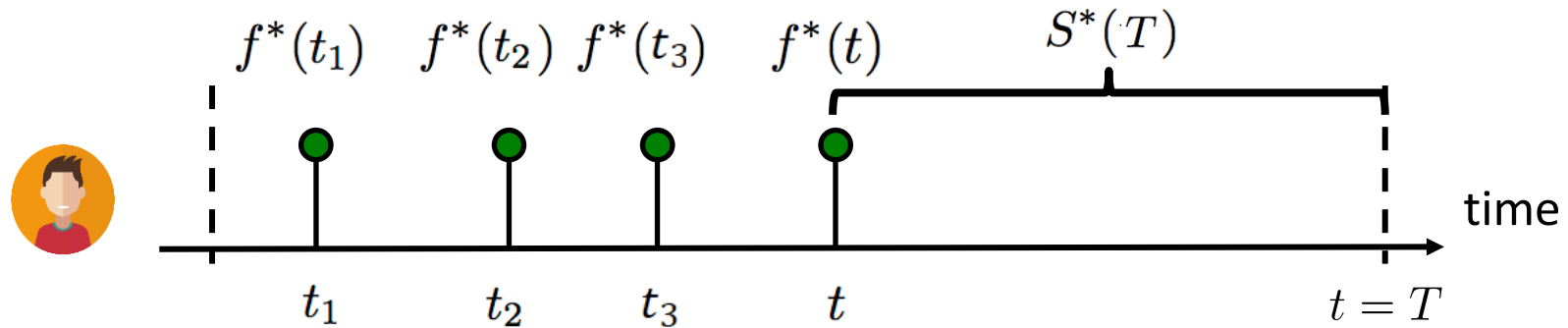
$$\lambda^*(t_1) \; \lambda^*(t_2) \; \lambda^*(t_3) \; \lambda^*(t) \; \exp\left(-\int_0^T \lambda^*(\tau)\,d\tau\right)$$

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

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

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

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

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

Suitable for model design and **interpretable**:

1. **Intensities only need to be nonnegative**

2. **Easy to combine timelines**

Relation between f*, F*, S*, λ*
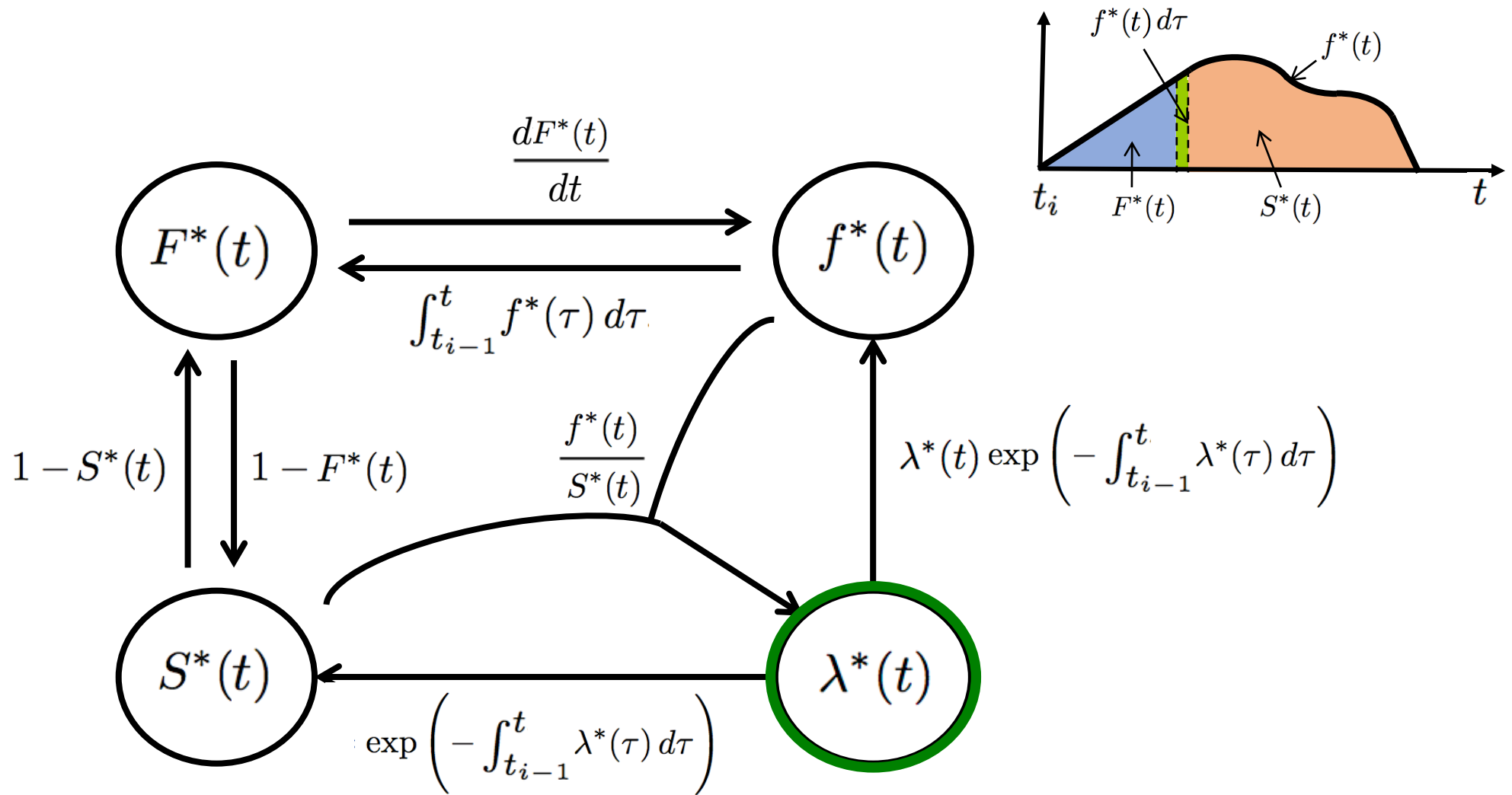
$$F^*(t) \quad \xrightarrow{\frac{dF^*(t)}{dt}} \quad f^*(t)$$

$$\int_{t_{i-1}}^{t} f^*(\tau)\, d\tau$$

$$1 - S^*(t) \qquad 1 - F^*(t)$$

$$\frac{f^*(t)}{S^*(t)}$$

$$\lambda^*(t) \exp\left(-\int_{t_{i-1}}^{t} \lambda^*(\tau)\, d\tau\right)$$

$$S^*(t) \qquad \exp\left(-\int_{t_{i-1}}^{t} \lambda^*(\tau)\, d\tau\right) \qquad \lambda^*(t)$$

Representation:
Temporal Point Processes

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

Poisson process
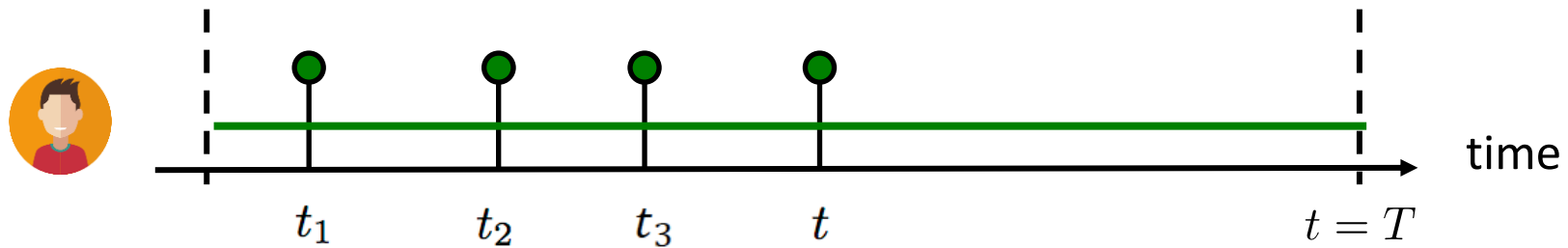
Intensity of a Poisson process

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

Observations:

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

Fitting & sampling from a Poisson

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

time

$t_1 \quad t_2 \quad t_3 \quad\quad\quad t = T$
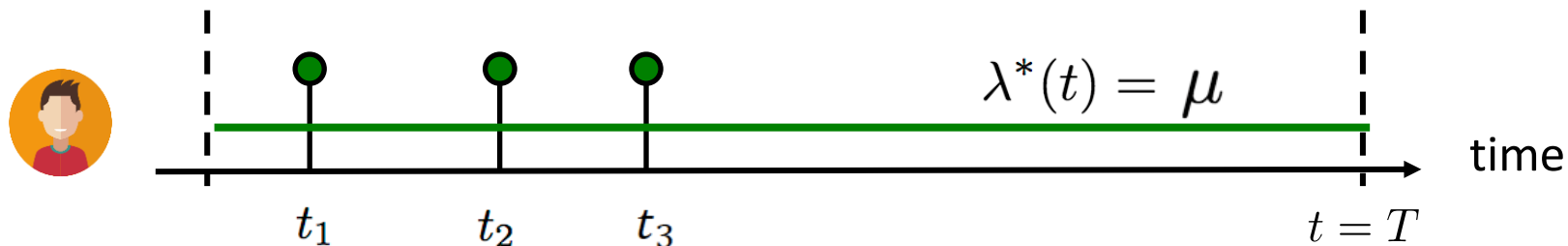
Fitting by maximum likelihood:

$$\mu^* = \underset{\mu}{\mathrm{argmax}}\ 3\log\mu\ -\ \mu\,T\ =\ \frac{3}{T}$$

Sampling using inversion sampling:

$$Uniform(0,1)$$

$$t \sim \mu\exp\left(-\mu(t - t_3)\right) \quad\Rightarrow\quad t = -\frac{1}{\mu}\log(1 - u) + t_3$$

$$\underbrace{\qquad\qquad\qquad}_{f_t^*(t)} \qquad\qquad\qquad \underbrace{\qquad\qquad\qquad\qquad}_{F_t^{-1}(u)}$$

Inhomogeneous Poisson process
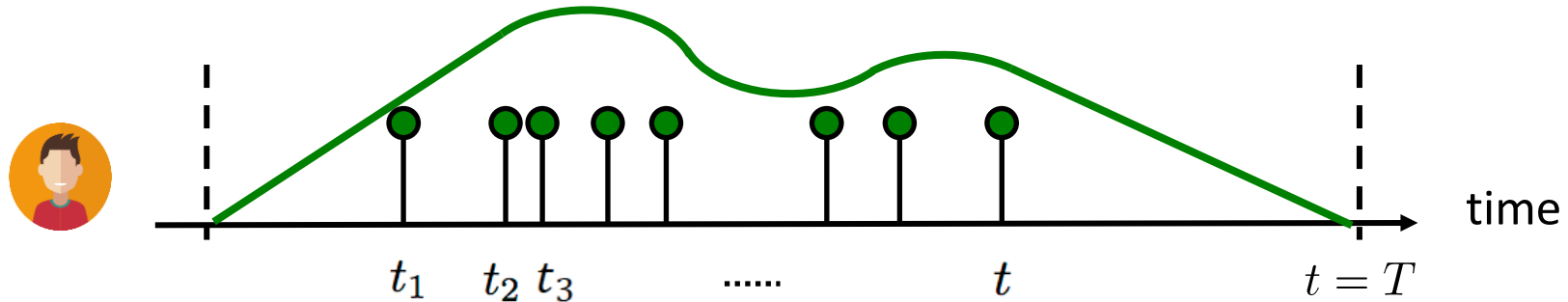
Intensity of an inhomogeneous Poisson process

$$\lambda^*(t) = g(t) \geqslant 0 \quad \textbf{(Independent of history)}$$

Example:

$$\lambda^*(t) = \sum_j \alpha_j \, k(t - t_j)$$
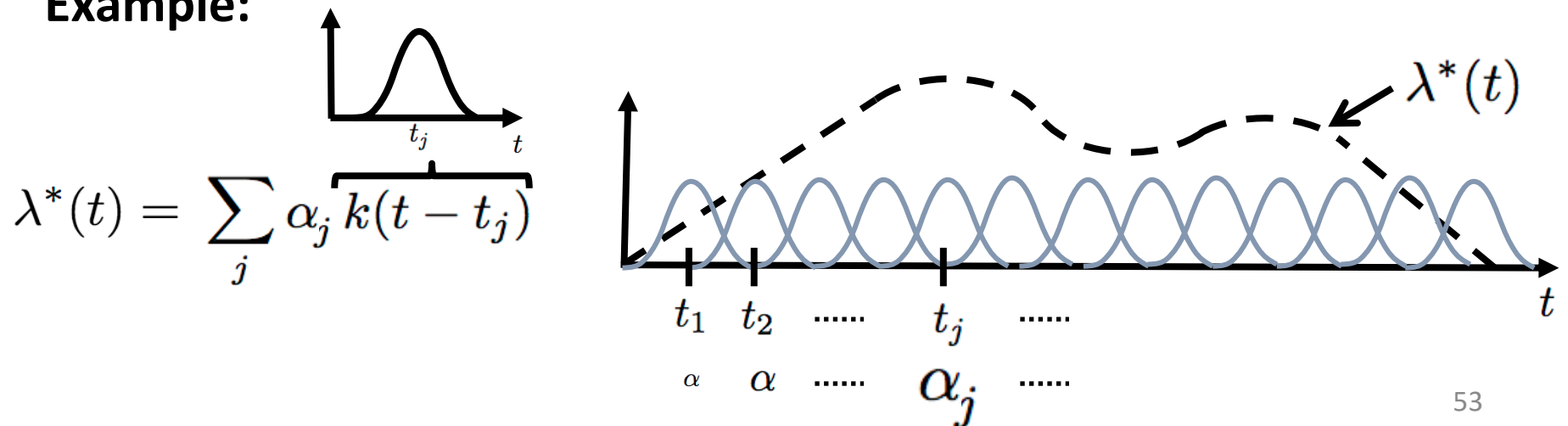
Fitting & sampling from inhomogeneous Poisson

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

time

$t_1 \quad t_2 \; t_3 \qquad \ldots\ldots \qquad t \qquad t = T$
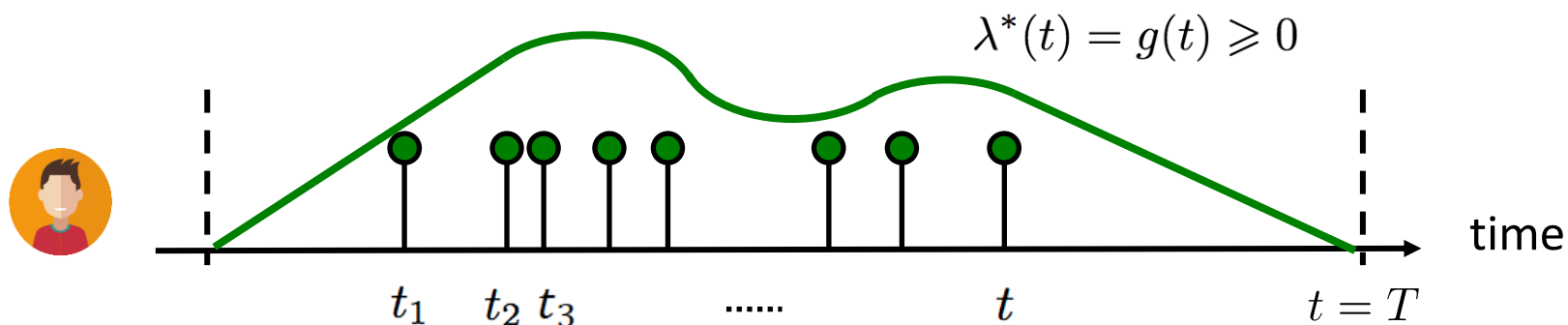
Fitting by maximum likelihood: $\displaystyle \operatorname*{maximize}_{g(t)} \; \sum_{i=1}^{n} \log g(t_i) - \int_0^T g(\tau) \, d\tau$

Sampling using thinning (reject. sampling) + inverse sampling:

1. Sample t from Poisson process with intensity μ using inverse sampling

2. Generate $u_2 \sim Uniform(0,1)$

3. Keep the sample if $u_2 \leq g(t) \big/ \mu$

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

54

Terminating (or survival) process
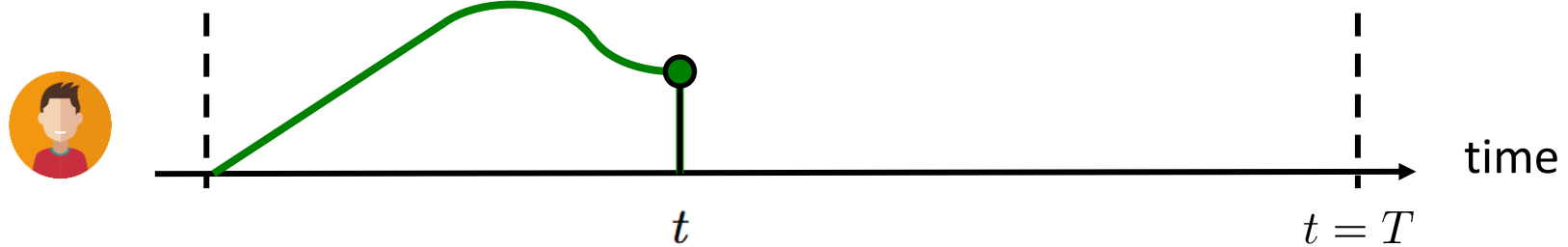
Intensity of a terminating (or survival) process

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

Observations:

1. **Limited number of occurrences**

Self-exciting (or Hawkes) process

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

Triggering kernel
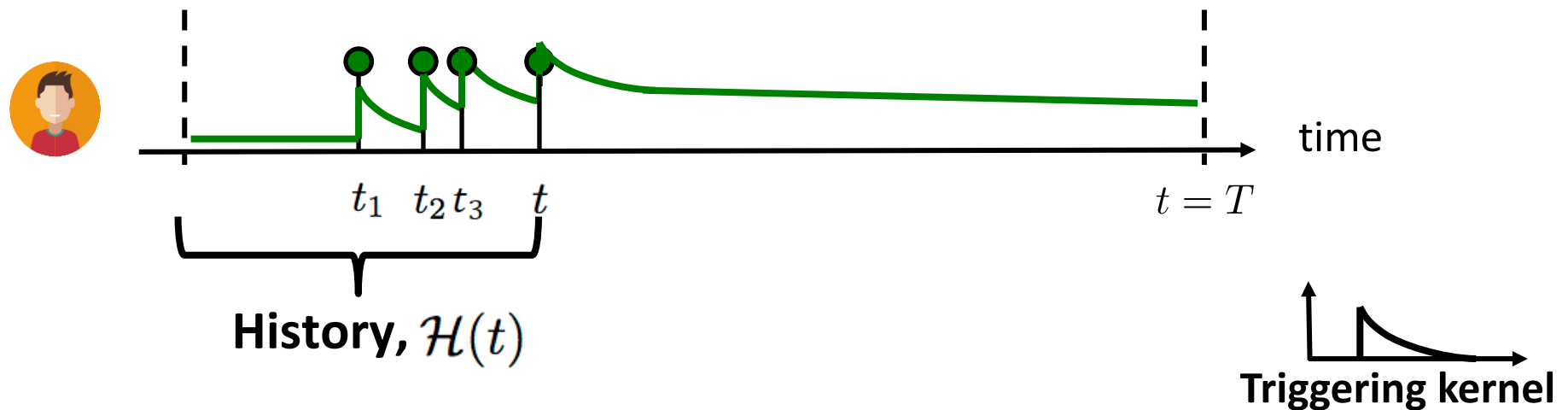
Intensity of self-exciting (or Hawkes) process:

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

$$= \mu + \alpha \kappa_\omega(t) \star dN(t)$$

Observations:

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

Fitting a Hawkes process from a recorded timeline

$$\lambda_0 = \lambda^*(t_3)$$

$t_1 \quad t_2\, t_3 \qquad\qquad\qquad t = T$
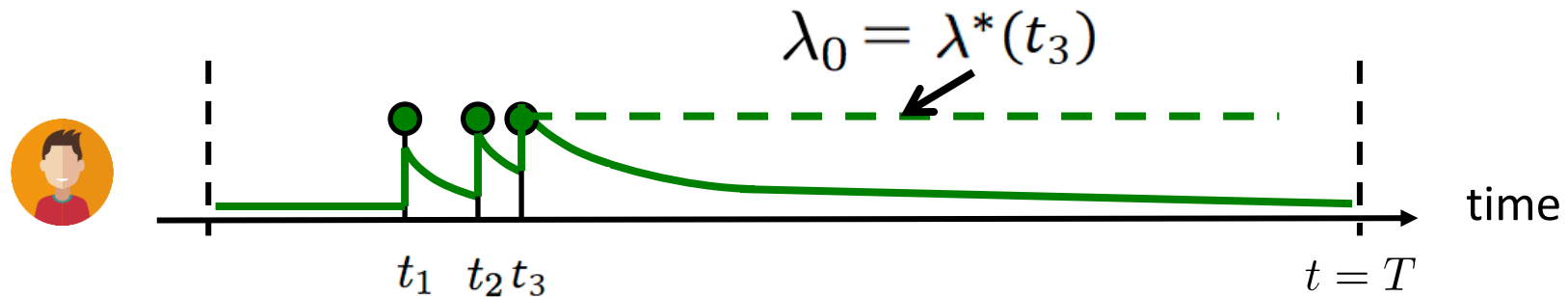
time

Fitting by maximum likelihood:

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

The max. likelihood
is **jointly convex**
in μ and α

Sampling using thinning (reject. sampling) + inverse sampling:

Key idea: the maximum of the intensity λ_0 changes over time

Summary

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

Poisson processes:

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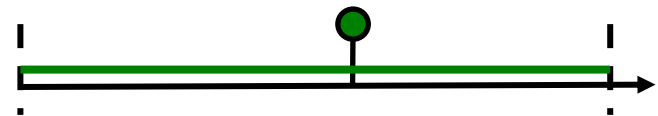
Inho

$$\text{We know } \textbf{how to fit} \text{ them}$$
$$\text{and } \textbf{how to sample} \text{ from them}$$

Term

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

Self-exciting point processes:

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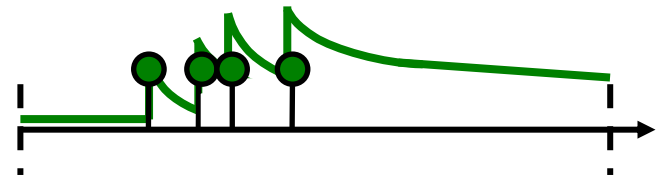
Representation:
Temporal Point Processes

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

Mutually exciting process
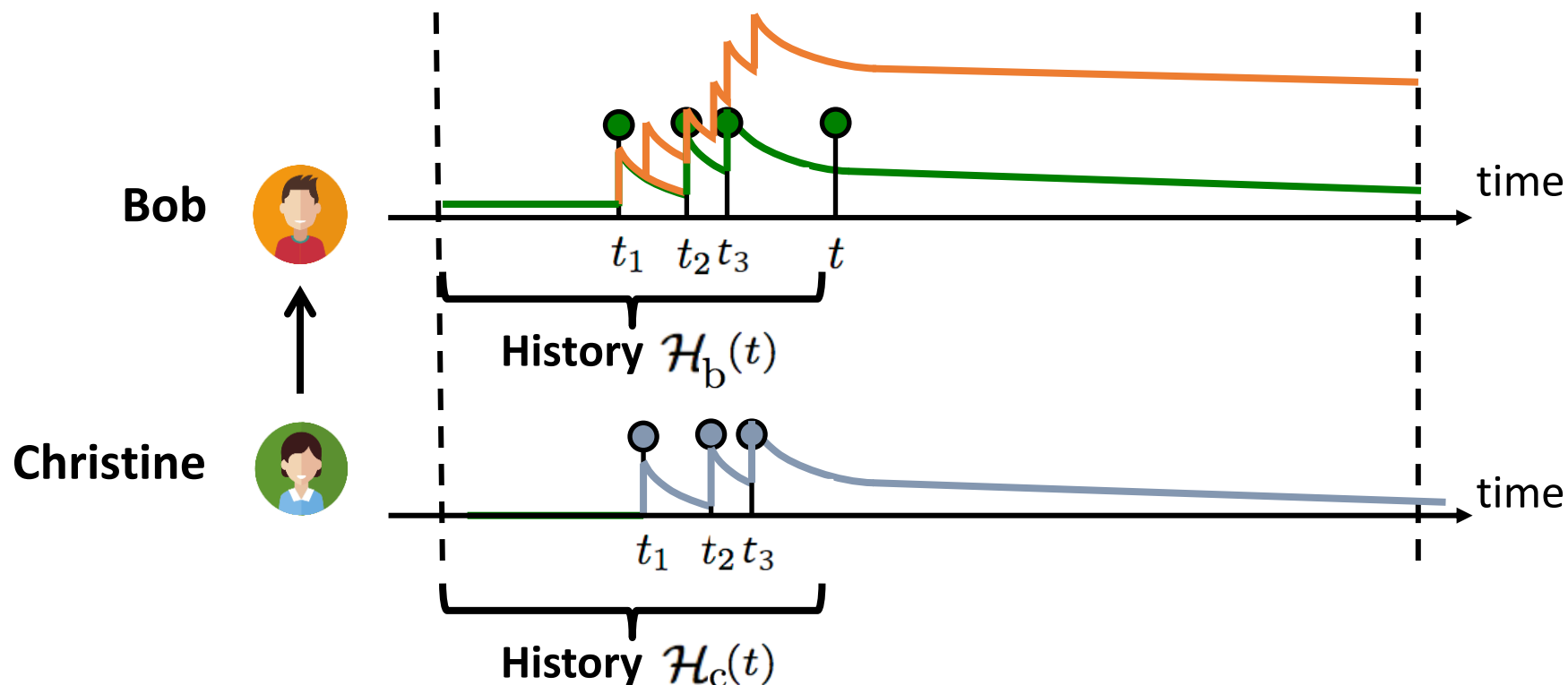
Clustered occurrence affected by neighbors

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

Mutually exciting terminating process
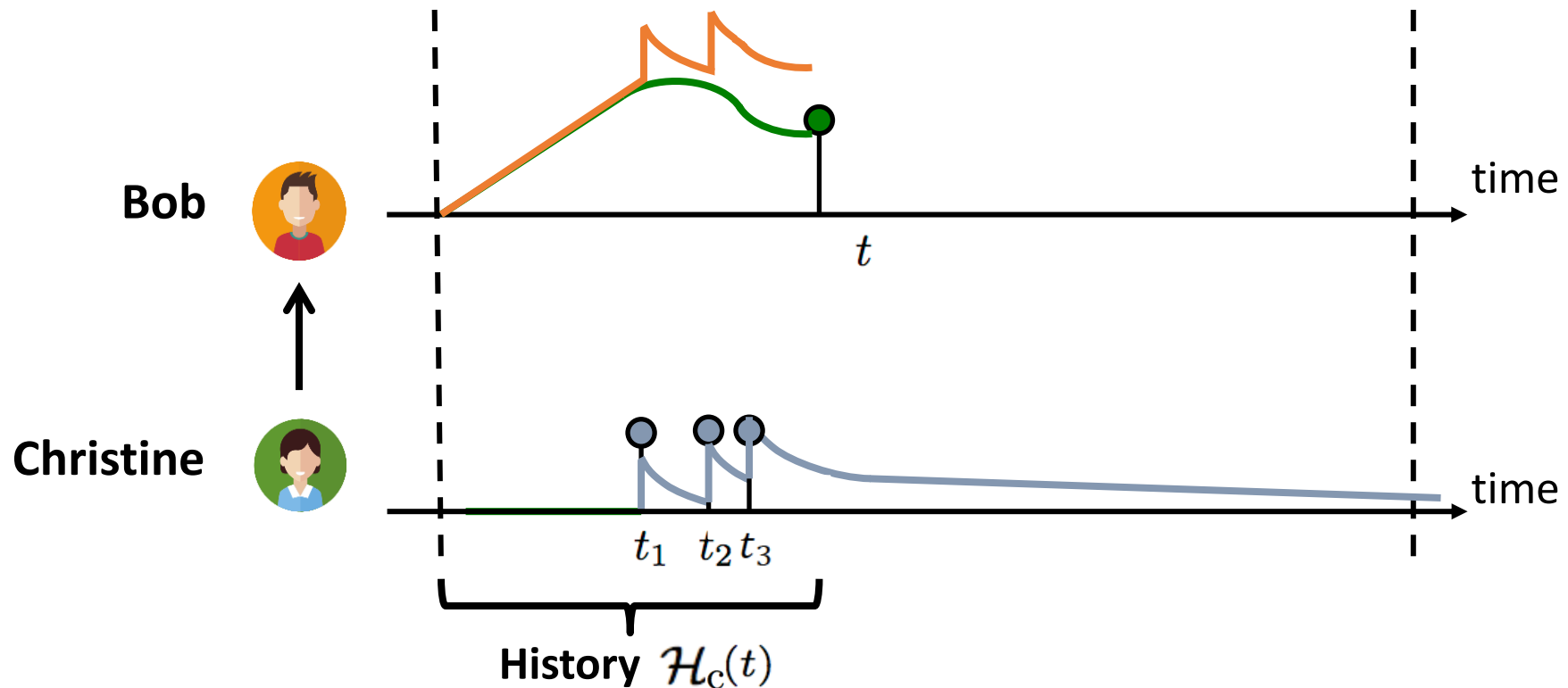
Clustered occurrence affected by neighbors

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

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**

$$N(t) \in \{0\} \cup \mathbb{Z}^+$$

$x(t)$

$y(t)$

History, $\mathcal{H}(t)$
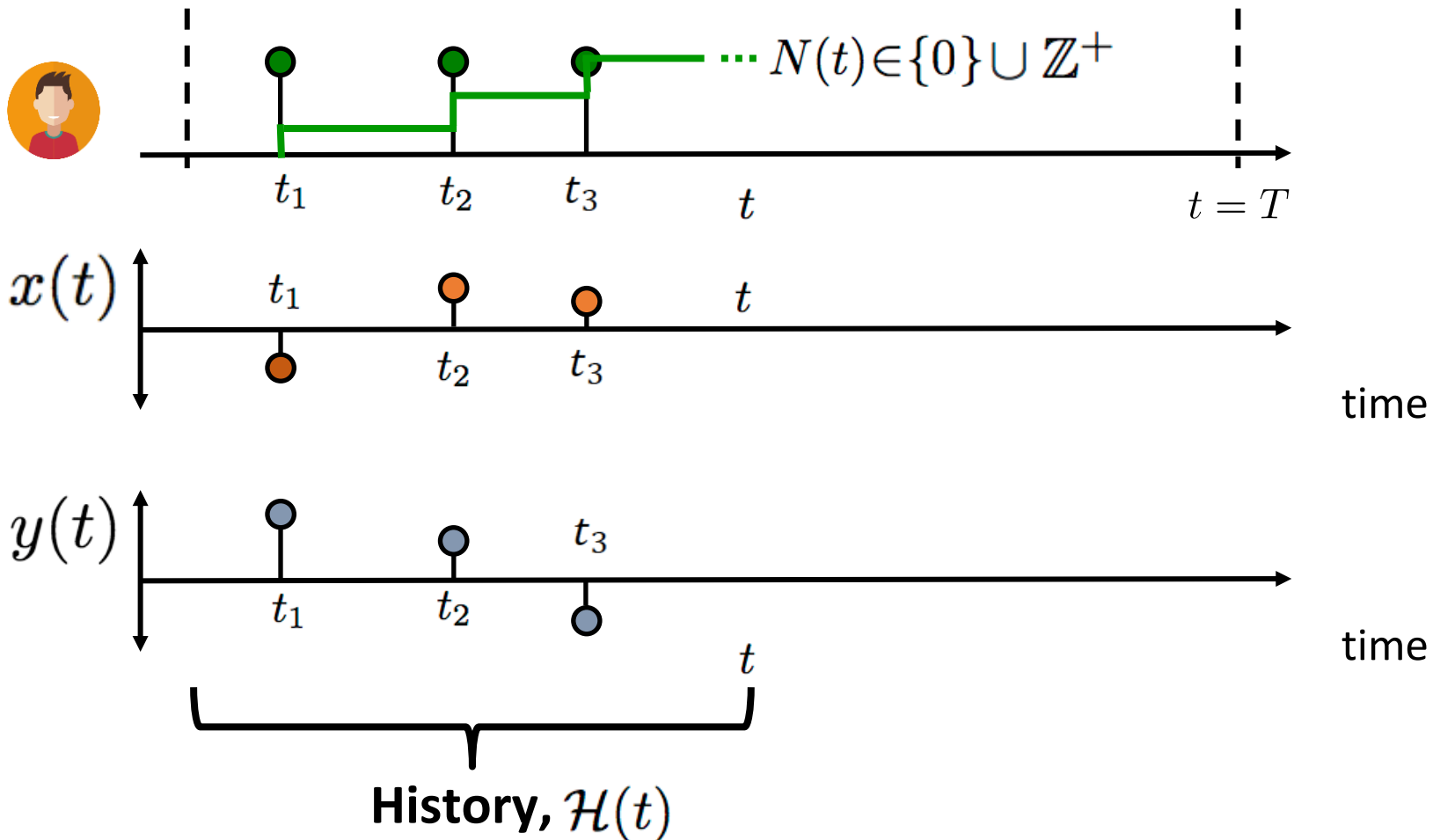
Independent identically distributed marks
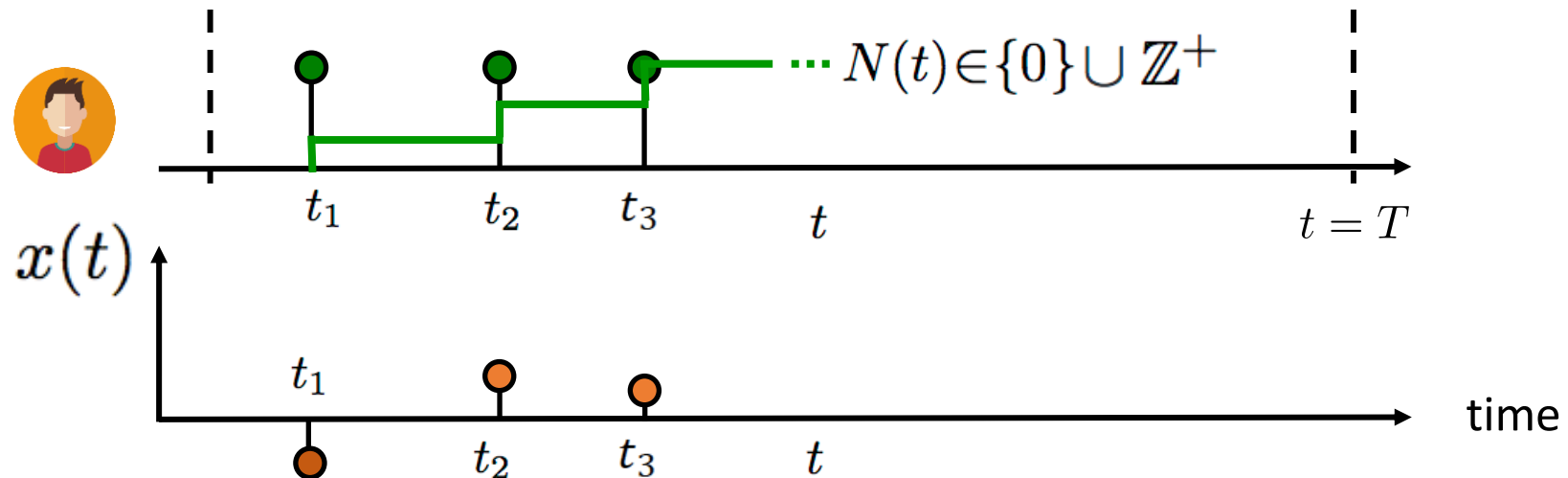
Distribution for the marks:

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

Observations:

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

Dependent marks: SDEs with jumps

$$N(t) \in \{0\} \cup \mathbb{Z}^+$$

History, $\mathcal{H}(t)$
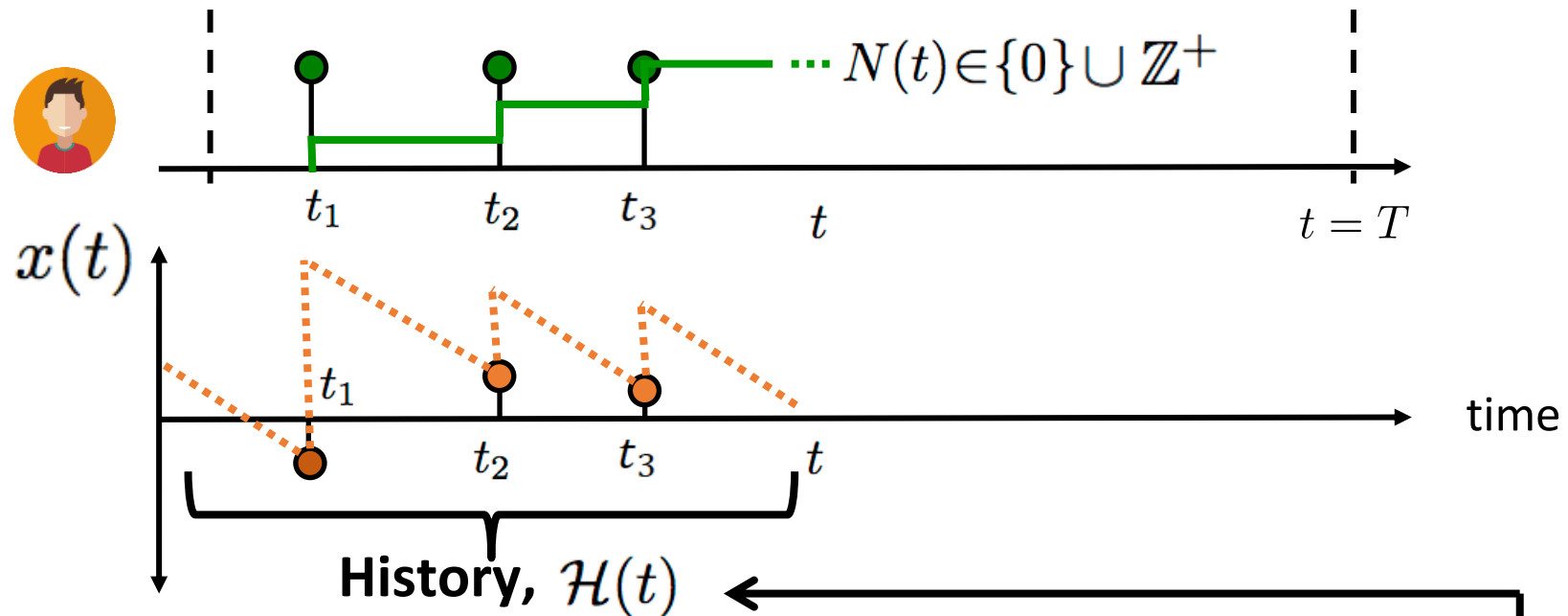
Marks given by stochastic differential equation with jumps:

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

Observations: **Drift**   **Event influence**

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

Dependent marks: distribution + SDE with jumps
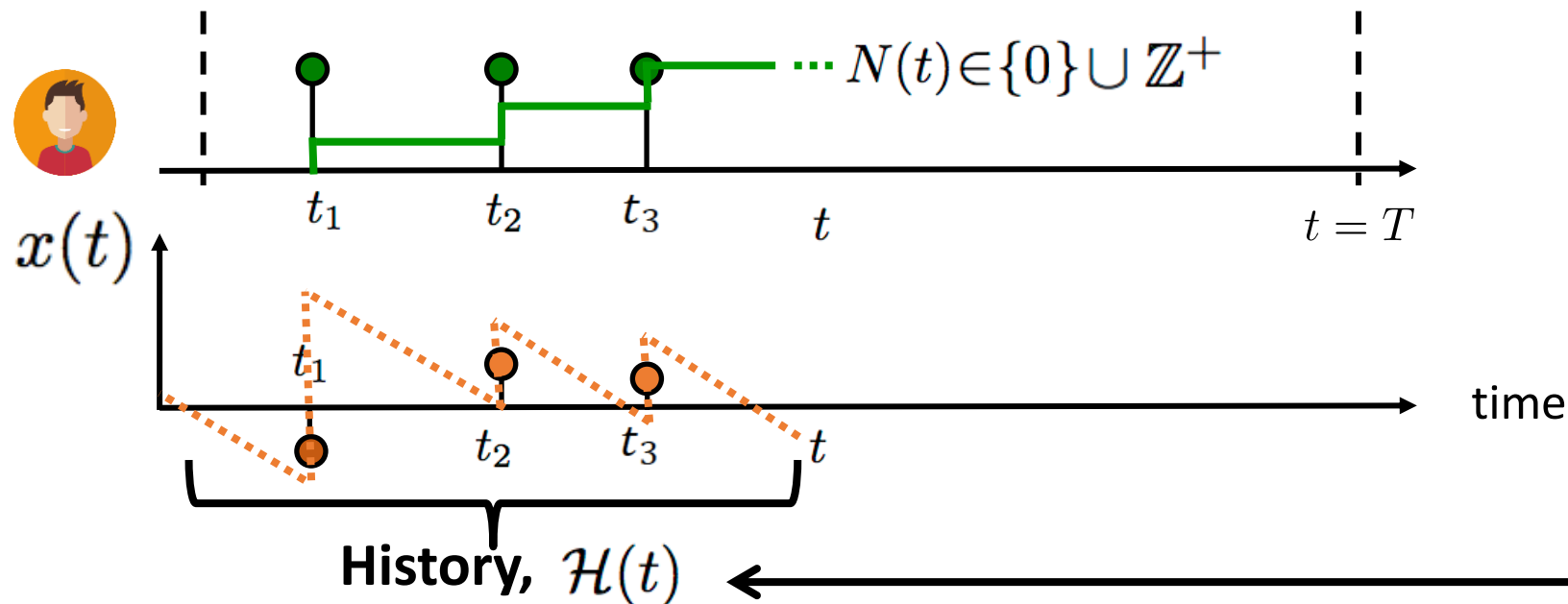
Distribution for the marks:

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

Observations:

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

Mutually exciting + marks

Bob

$N(t) \in \{0\} \cup \mathbb{Z}^+$

$x(t)$

time

Christine

$M(t) \in \{0\} \cup \mathbb{Z}^+$
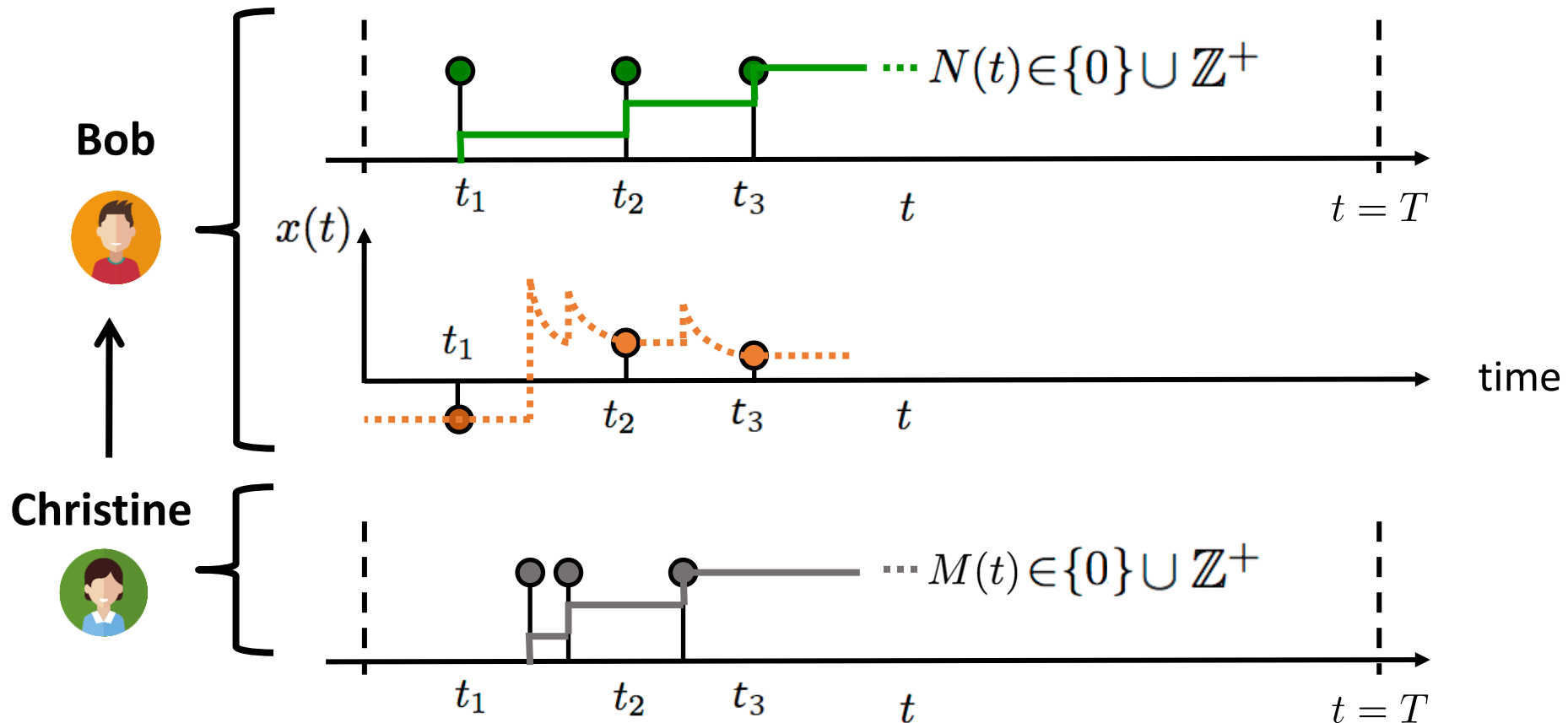
Marks affected by neighbors

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

Marked TPPs as stochastic dynamical systems

Example: Susceptible-Infected-Susceptible (SIS)
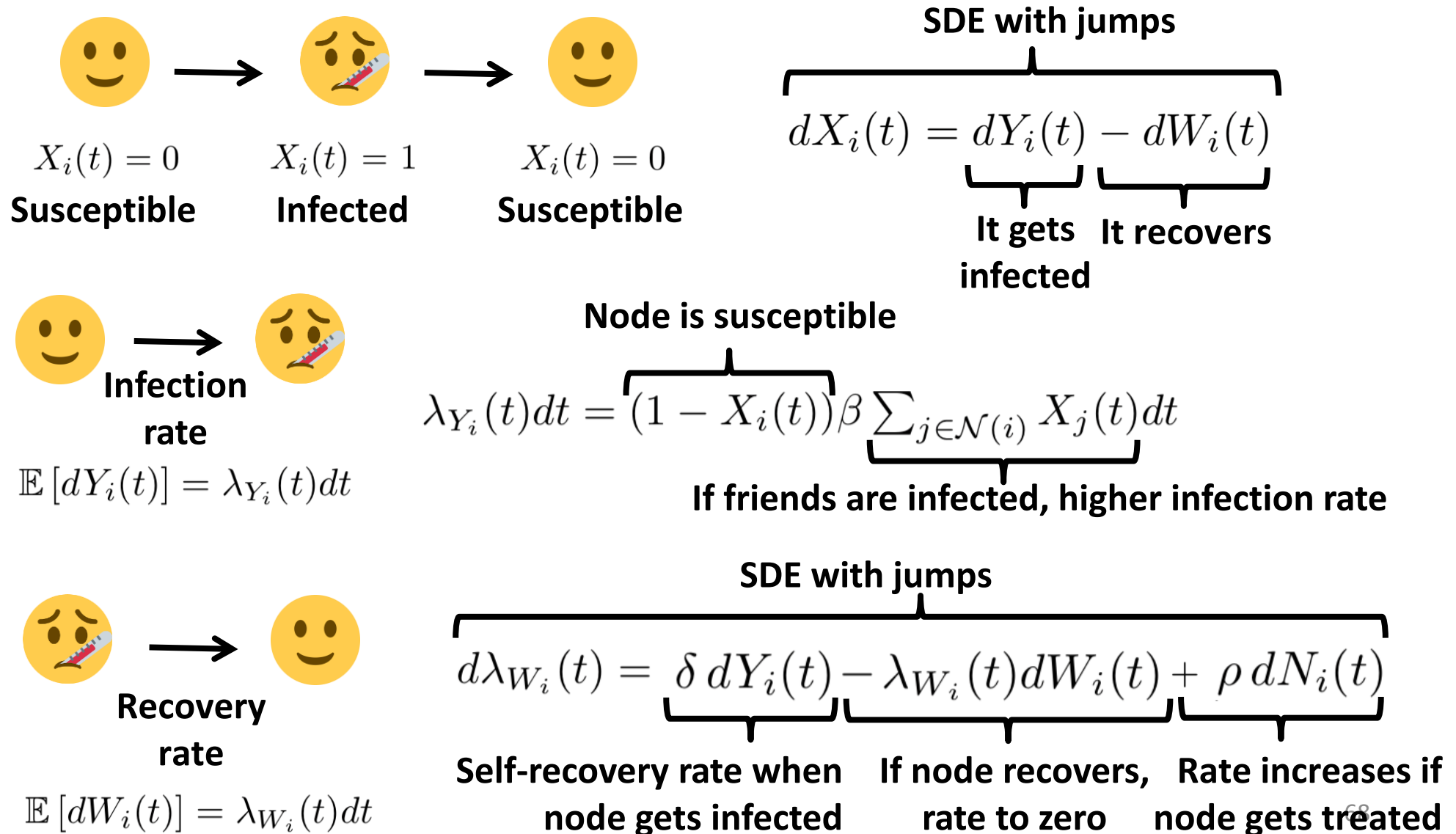
$X_i(t) = 0$
Susceptible

$X_i(t) = 1$
Infected

$X_i(t) = 0$
Susceptible

SDE with jumps

$$dX_i(t) = dY_i(t) - dW_i(t)$$

It gets infected   **It recovers**

Infection rate

$$\mathbb{E}[dY_i(t)] = \lambda_{Y_i}(t)dt$$

Node is susceptible

$$\lambda_{Y_i}(t)dt = (1 - X_i(t))\beta \sum_{j \in \mathcal{N}(i)} X_j(t)dt$$

If friends are infected, higher infection rate

Recovery rate

$$\mathbb{E}[dW_i(t)] = \lambda_{W_i}(t)dt$$

SDE with jumps

$$d\lambda_{W_i}(t) = \delta\, dY_i(t) - \lambda_{W_i}(t)dW_i(t) + \rho\, dN_i(t)$$

Self-recovery rate when node gets infected   **If node recovers, rate to zero**   **Rate increases if node gets treated**

Models & Inference

1. Modeling event sequences
2. Clustering event sequences
3. Capturing complex dynamics
4. Causal reasoning on event sequences

Event sequences as cascades

S → D means D follows S

Christine — 3.00pm
Bob — 3.25pm
Beth — 3.27pm
Joe
David — 4.15pm
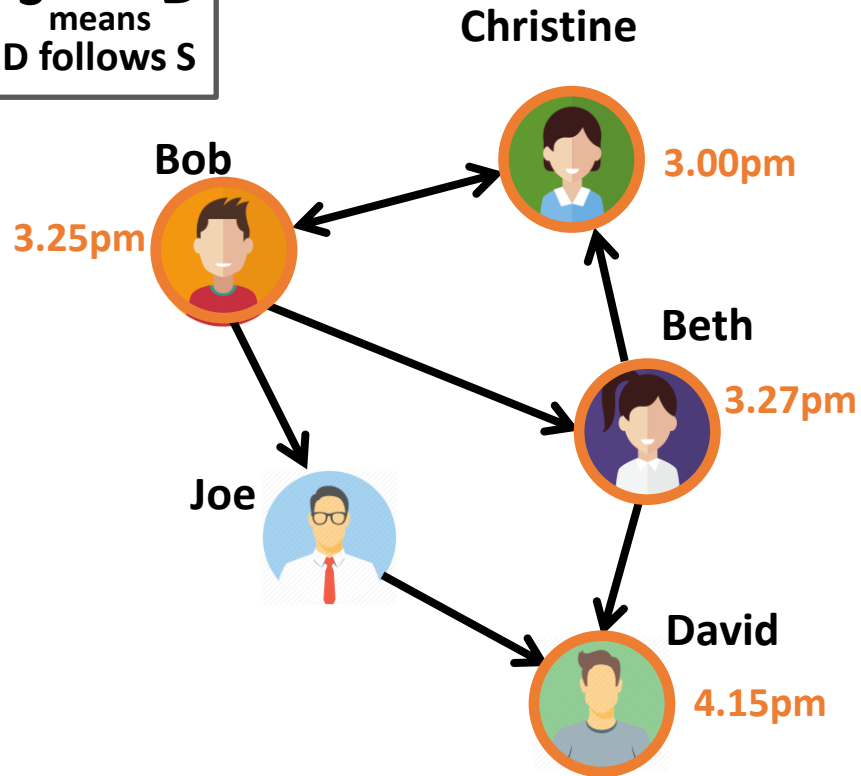
Information Diffusion

lipstick on a pig

i will reach out my hand to anyone to help me get this country moving again

our entire economy is in danger

i guess a small-town mayor is sort of like a community organizer except that you have actual responsibilities

effort to protect the american economy must not fail

decent person and a person that you do not have to be scared of as president of the united states

we have been blessed with five wonderful children who we love with all our heart and mean everything to us

the most serious financial crisis since the great depression

this is something that all of us will swallow hard and go forward with

all the parts of the internet are on the iphone

no way no how no mccain. barack obama is my candidate

fundamentals of our economy are strong

who is the real barack obama

i think when you spread the wealth around it's good for everybody

answering that question with specificity is above my pay grade

president's job to deal with more than one thing at once

he's palling around with terrorists

i am not president bush

he doesn't look like all those other presidents on the dollar bills

hey can i call you joe

she is a diva she takes no advice from anyone

i think i'll have my staff get to you

russian aggression must not go unanswered

8/1 8/8 8/15 8/22 8/29 9/5 9/12 9/19 9/26 10/3 10/10 10/17 10/24 10/31

800 700 600 500 400 300 200

[Leskovec et al., 2009]

Disease Diffusion

Event: (t_i, u_i)

Time   User

Population state

Events

t_1 t_2 t_3 t_4 t_5 time

[Rizoiu et al., 2018]

An example: idea adoption

S ⟶ D means D follows S

Bob — 3.25pm
Christine — 3.00pm
Beth — 3.27pm
Joe
David — 4.15pm

They can have an impact in the off-line world

Friggeri et al., 2014

theguardian

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

Infection cascade representation

We represent an infection cascade using **terminating temporal point processes**:
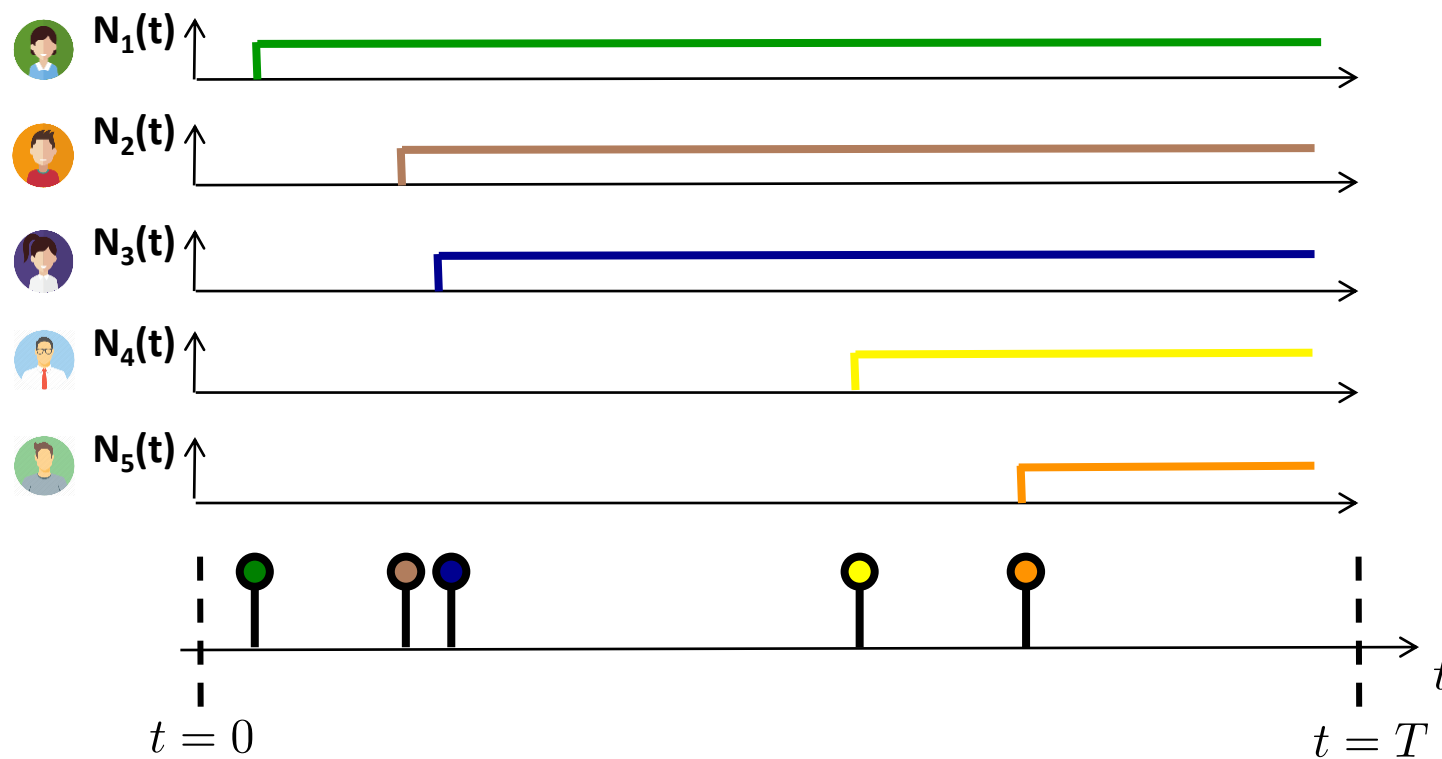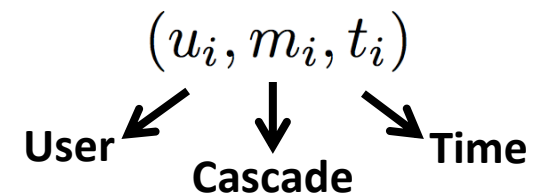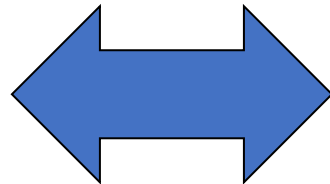
Infection intensity
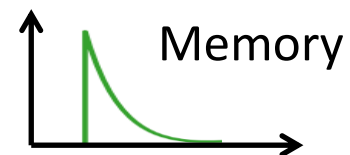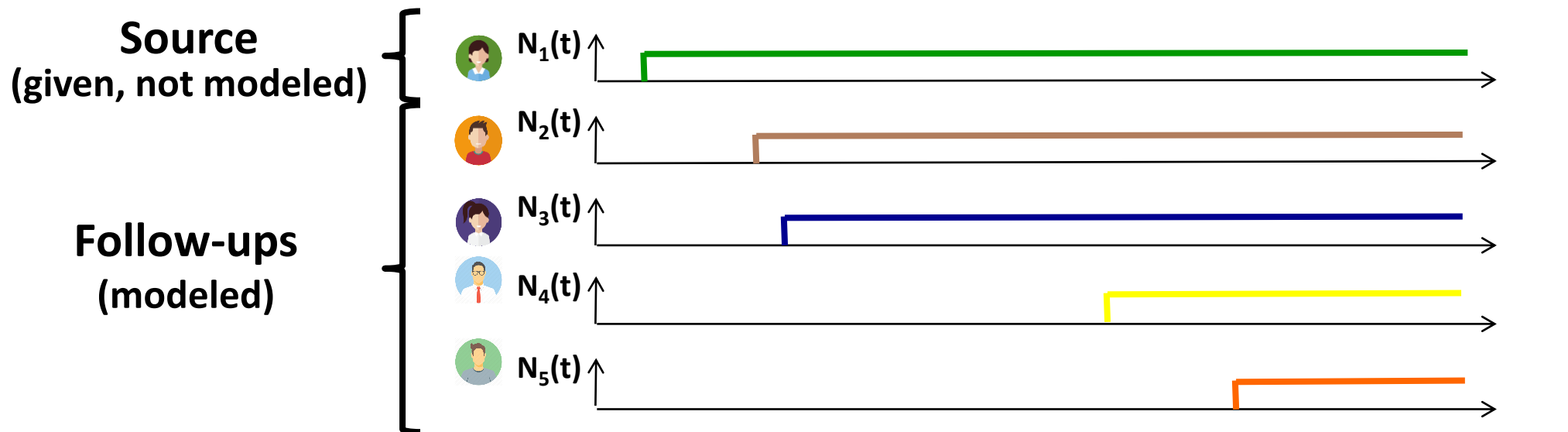
$$\lambda_u^*(t) = \underbrace{(1 - N_u(t))}_{\substack{\textbf{Users get} \\ \textbf{infected only} \\ \textbf{once}}} \sum_{v \in [m]} \underbrace{b_{vu}}_{\substack{\textbf{Influence from} \\ \textbf{user v on user u}}} \underbrace{\sum_{e_i \in \mathcal{H}_v(t)} \kappa(t - t_i)}_{\substack{\textbf{Previous} \\ \textbf{infections of user v}}}$$

[Gomez-Rodriguez et al., ICML 2011]

Model inference from multiple cascades

Conditional intensities

$\lambda_u^*(t)$

Diffusion log-likelihood

$$\mathcal{L} = \sum_{u=1}^{n} \log \lambda_u^*(t_u) - \int_0^T \lambda_u^*(\tau)\, d\tau$$

Maximum likelihood approach to find model parameters!

Sum up log-likelihoods of multiple cascades!

Theorem. For any choice of parametric memory, the **maximum likelihood** problem is **convex**.

[Gomez-Rodriguez et al., ICML 2011]

In some cases, influence change over time:

#greece retweets

Propagation over networks with variable influence

0 ———————————— T

[Gomez-Rodriguez et al., WSDM 2013]
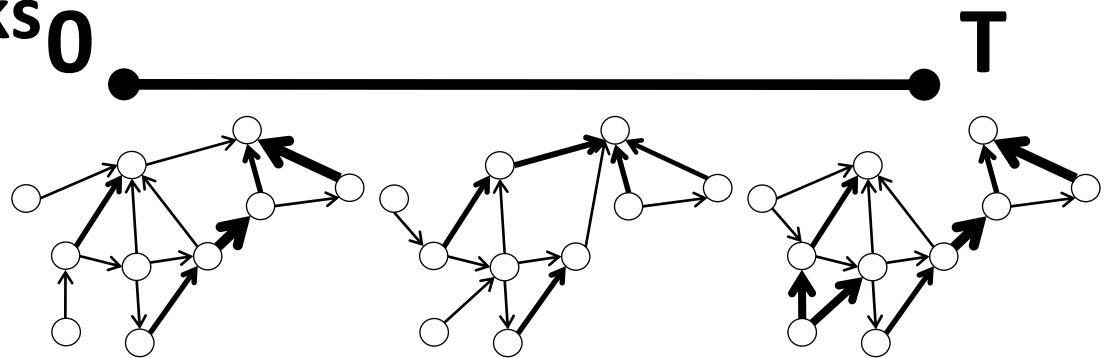
Recurrent events: beyond cascades

Up to this point, each users is only infected once, and event sequences can be seen as cascades.

In general, users perform recurrent events over time. E.g., people repeatedly express their opinion online:
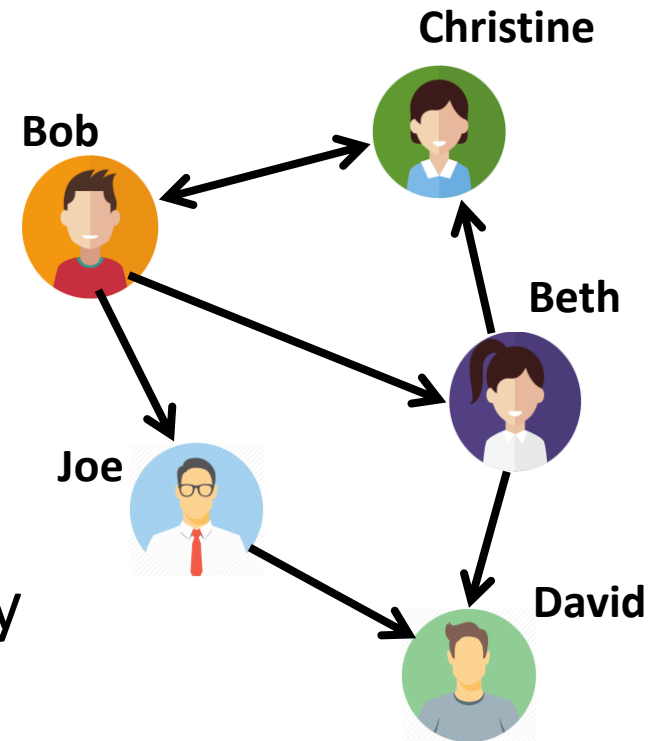
CNN: How social media is revolutionizing debates

The New York Times
Social Media Are Giving a Voice to Taste Buds

TC
Twitter Unveils A New Set Of Brand-Centric Analytics

The New York Times
Campaigns Use Social Media to Lure Younger Voters

Recurrent events representation

We represent messages using **nonterminating temporal point processes**:

Recurrent event:

$$(u_i, t_i)$$

User   **Time**

[Farajtabar et al., NIPS 2014]
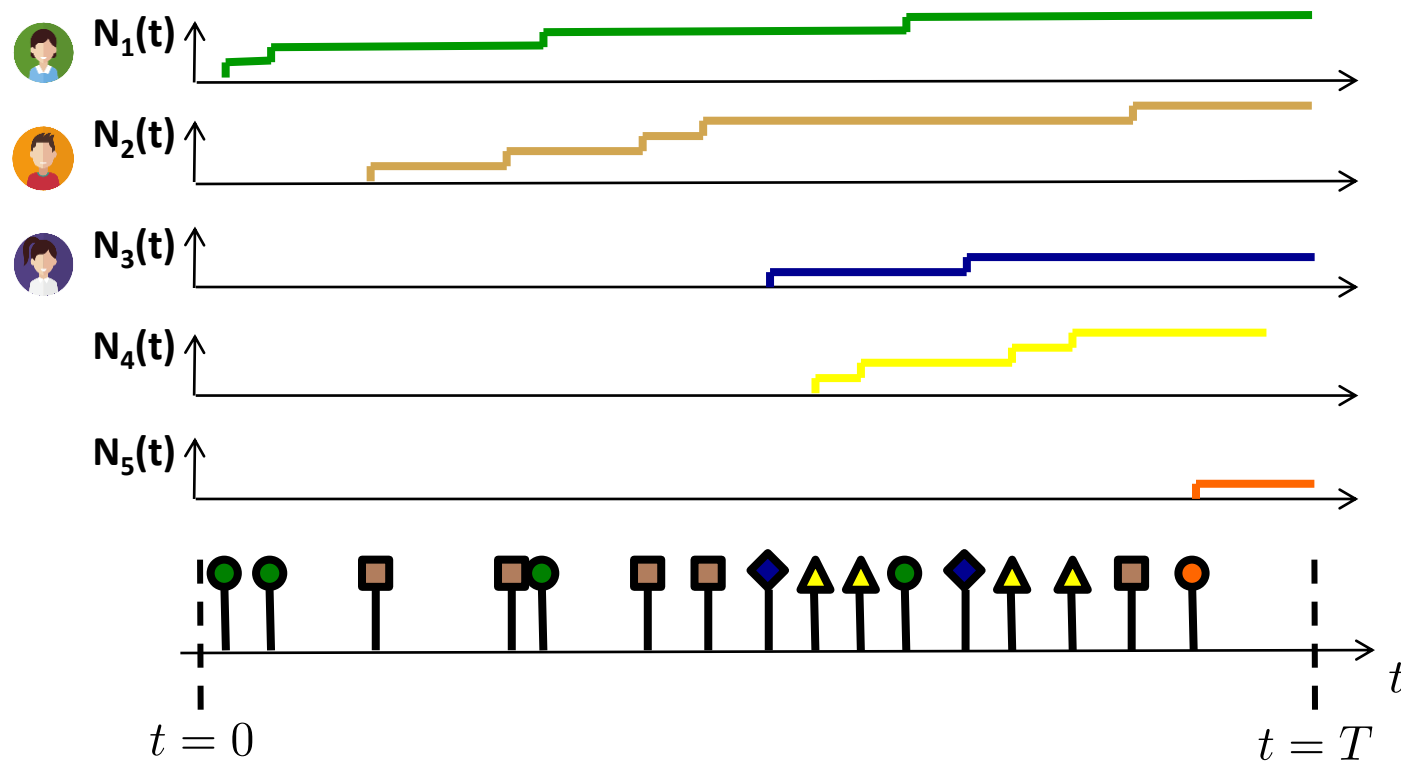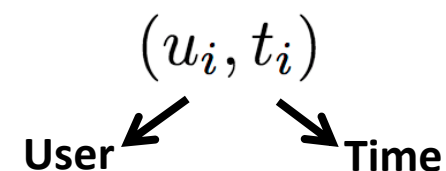
Recurrent events intensity
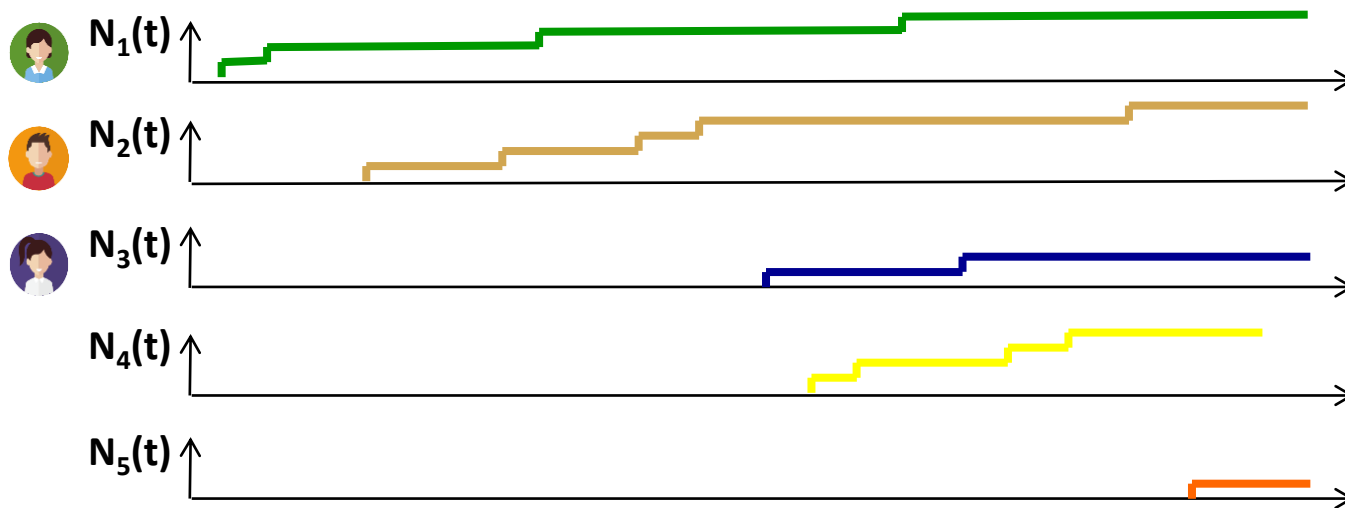
$$\underbrace{\lambda_u^*(t)}_{\substack{\text{User's} \\ \text{intensity}}} = \underbrace{\mu_u}_{\substack{\text{Events on her} \\ \text{own initiative}}} + \sum_{v \in [m]} \underbrace{b_{vu}}_{\substack{\text{Influence from} \\ \text{user v on user u}}} \underbrace{\sum_{e_i \in \mathcal{H}_v(t)} \kappa(t - t_i)}_{\substack{\text{Previous} \\ \text{messages by user v}}}$$
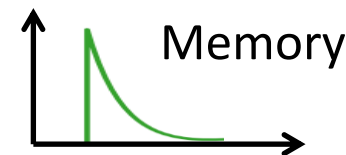
Cascade sources!

Memory

Hawkes process

[De et al., NIPS 2016]

Models & Inference

1. Modeling event sequences
2. Clustering event sequences
3. Capturing complex dynamics
4. Causal reasoning on event sequences

Event sequences

So far, we have assumed the cascade (topic, meme, etc.) that each event belongs to was known.
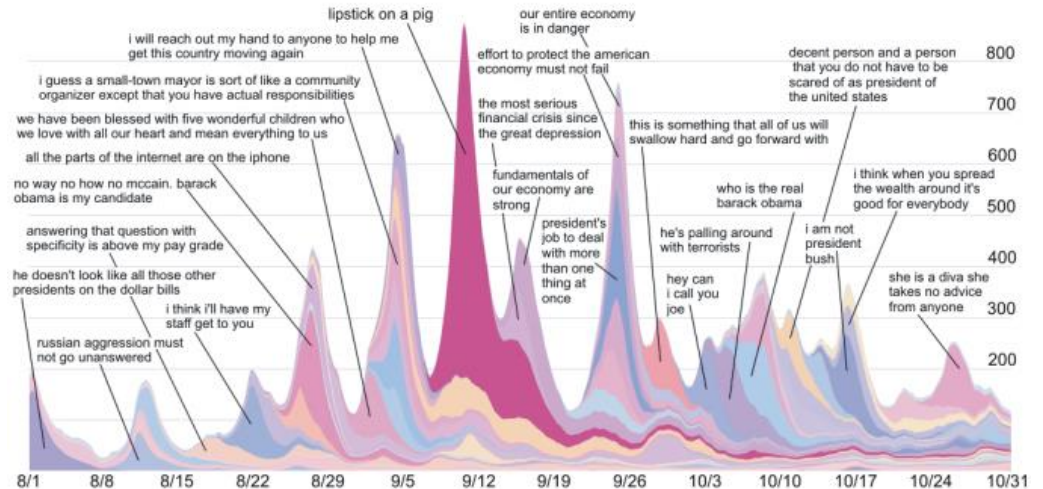
Often, the cluster (topic, meme, etc.) that each event in a sequence belongs to is not known:
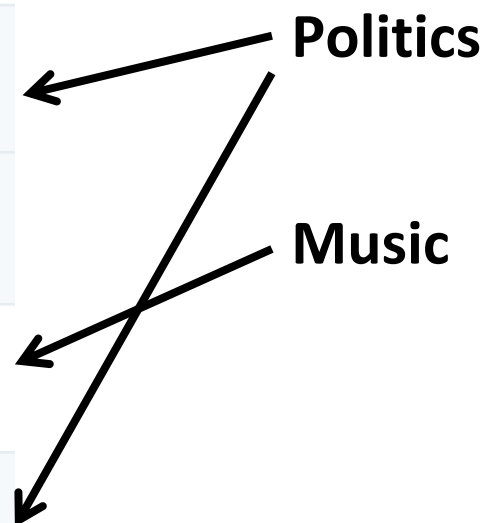
Assume the event **cluster to be hidden** and aim to automatically **learn the cluster assigments** from the data:
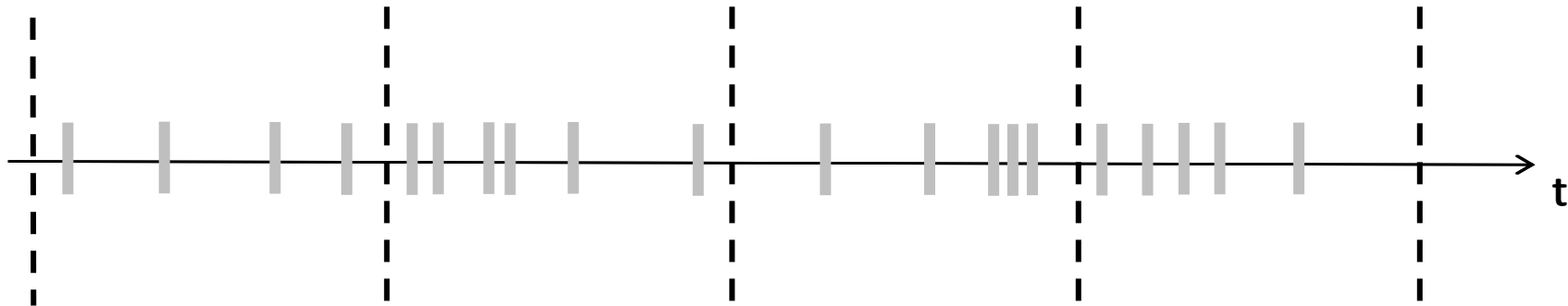
Bayesian methods to cluster event sequences in the context of:

Learning

Online News
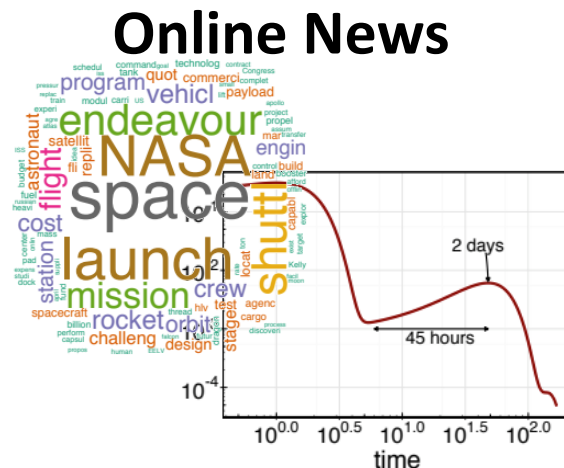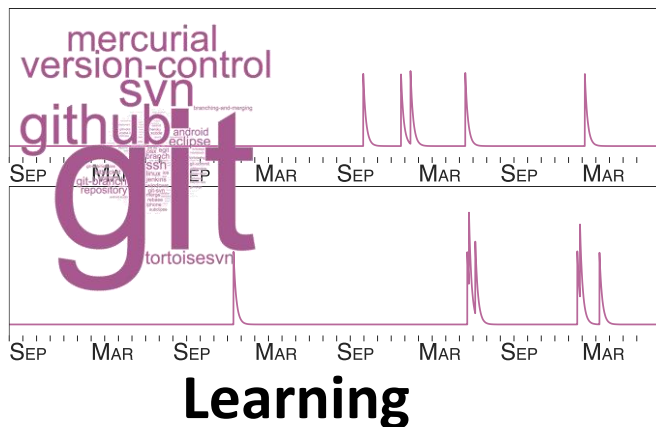
Health care

Method	DMHP
ICU Patient	**0.3778**
IPTV User	**0.2004**

[Du et al., 2015; Mavroforakis et al., 2017; Xu & Zha, 2017]
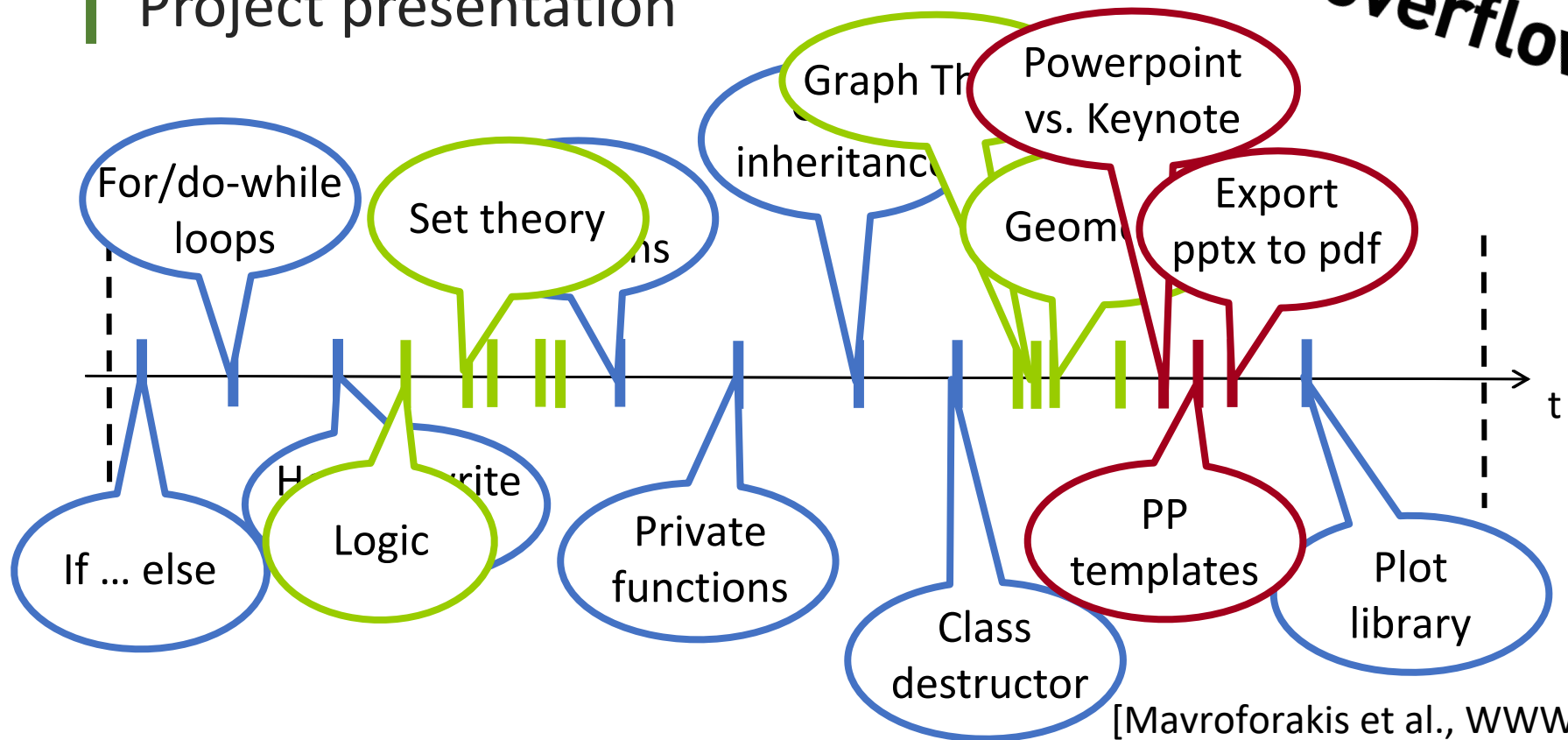
Events representation

We represent the events using **marked temporal point processes**:

$N_{u,\ell}(t)$

Task

Task

$t = 0$

$t = T$
$N_{u,\ell}(T) = 9$

t

Event: (t_n, p_n, q_n)

Time

Cluster
(hidden)

Content

[Mavroforakis et al., WWW 2017]

Cluster intensity

$$\lambda_{u,\ell}^{*}(t) = \mu_u \pi_\ell + \sum_{j:t_j \in \mathcal{H}_{u,\ell}(t)} k_{\theta_\ell}(t - t_j)$$

New cascade rate

Cluster popularity

Memory

Intensity or rate (events / hour)

Own initiative

Follow-up

Hawkes process

[Mavroforakis et al., WWW 2017]
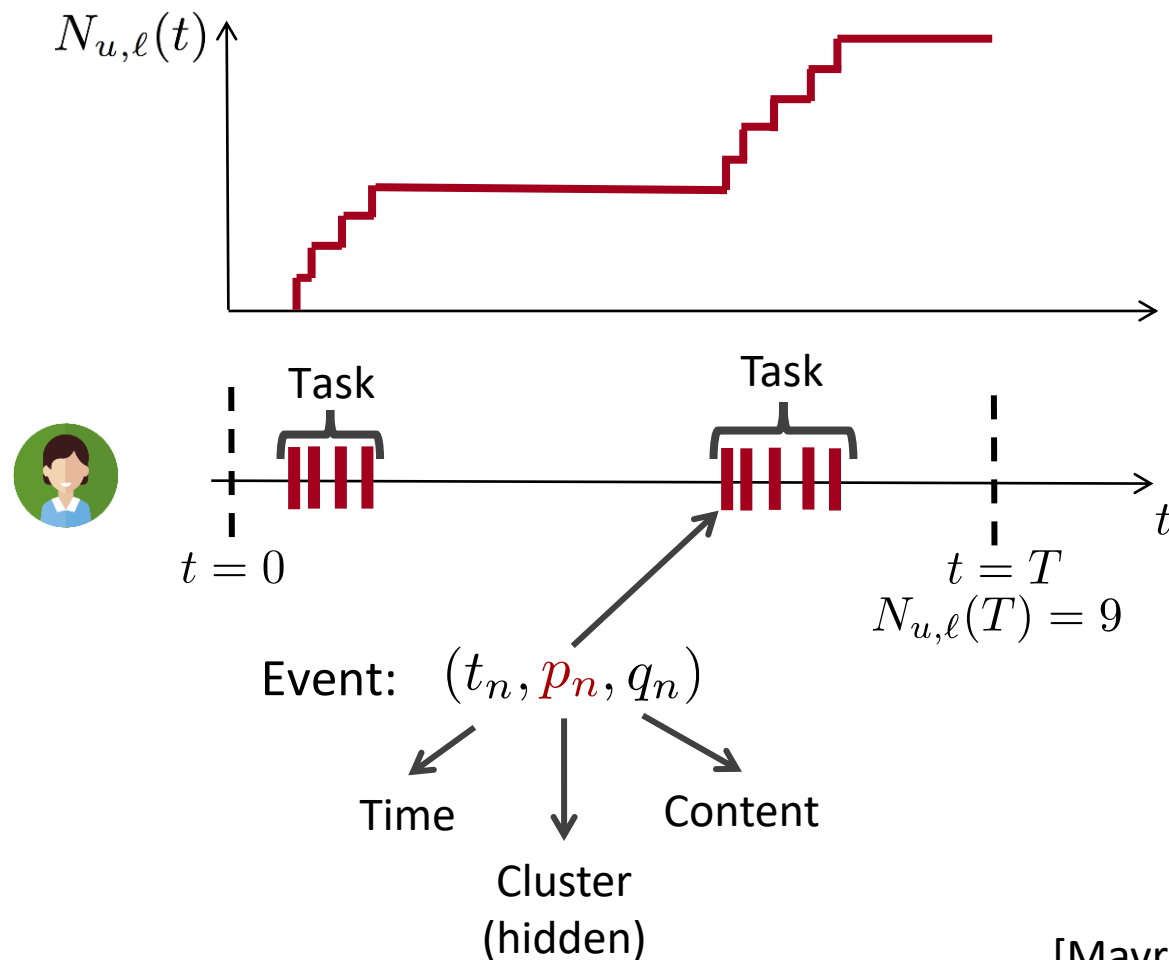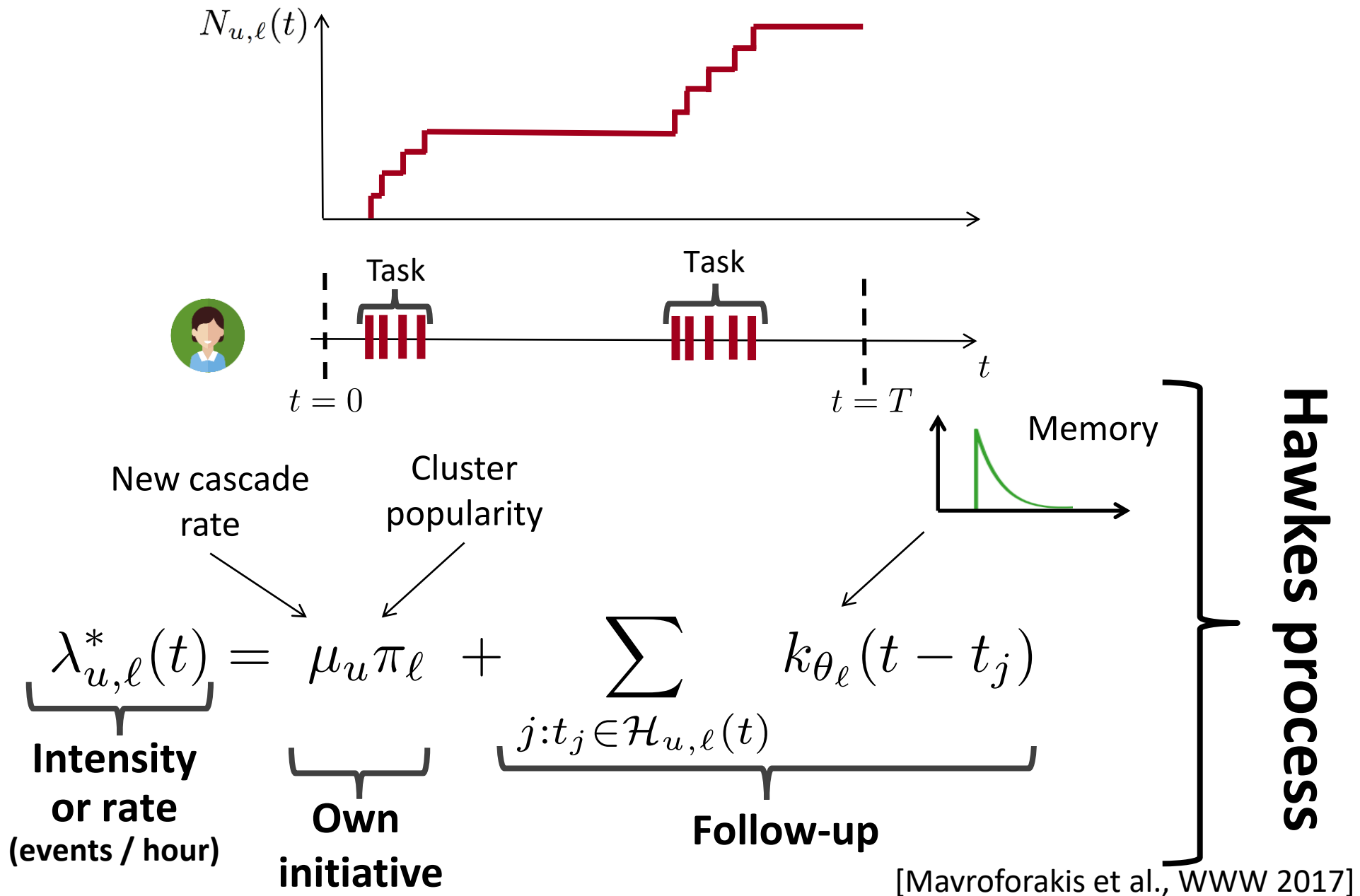
User events intensity

Users adopt more than one cluster:
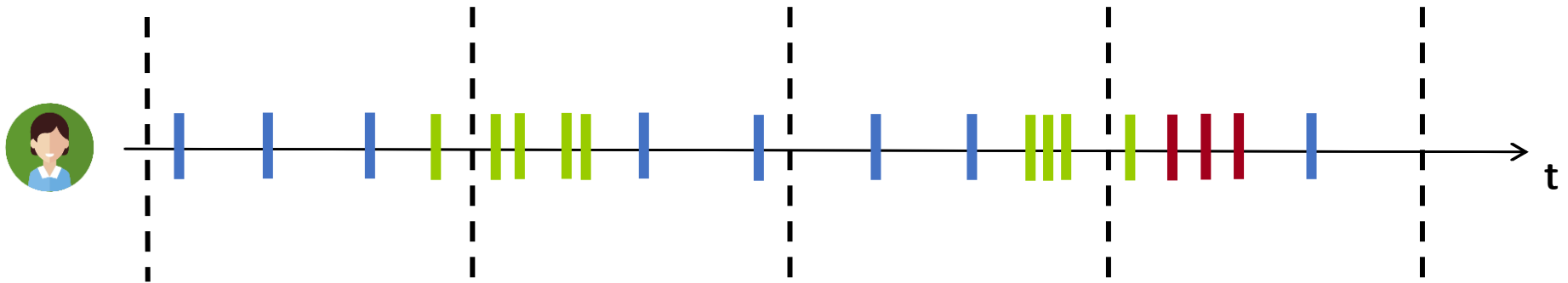
A user's learning events as a multidimensional Hawkes:

$$
\text{Time} \quad \text{cluster}
$$

$$
(t_n, p_n) \sim Hawkes \begin{pmatrix} \lambda_{u,1}^*(t) \\ \vdots \\ \lambda_{u,\infty}^*(t) \end{pmatrix}
$$

$$
\text{Content} \rightarrow q_n = \boldsymbol{\omega} \qquad \omega_j \sim Multinomial(\boldsymbol{\theta}_p)
$$

[Mavroforakis et al., WWW 2017]

People share same clusters

Different users adopt same clusters

Clus

-
- Shared parameters across users.

Efficient model inference using Sequential Montecarlo!

Details in the reference below!

[Mavroforakis et al., WWW 2017]

Content
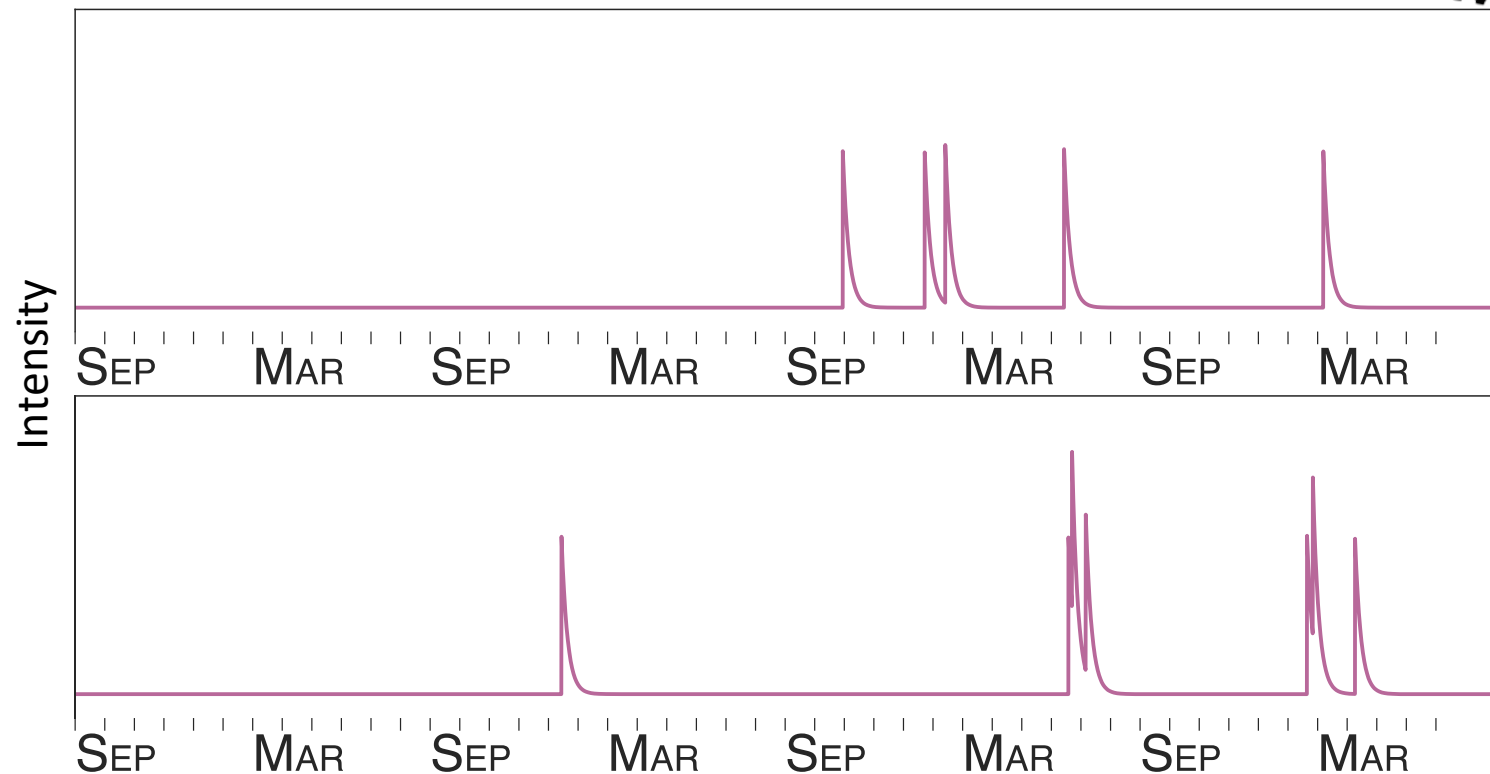
Intensities

Version control tasks tend to be specific, quickly solved after performing few questions

[Mavroforakis et al., WWW 2017]

Content
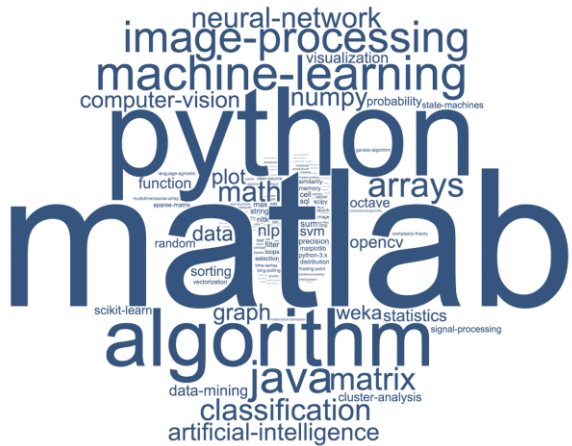
Intensities
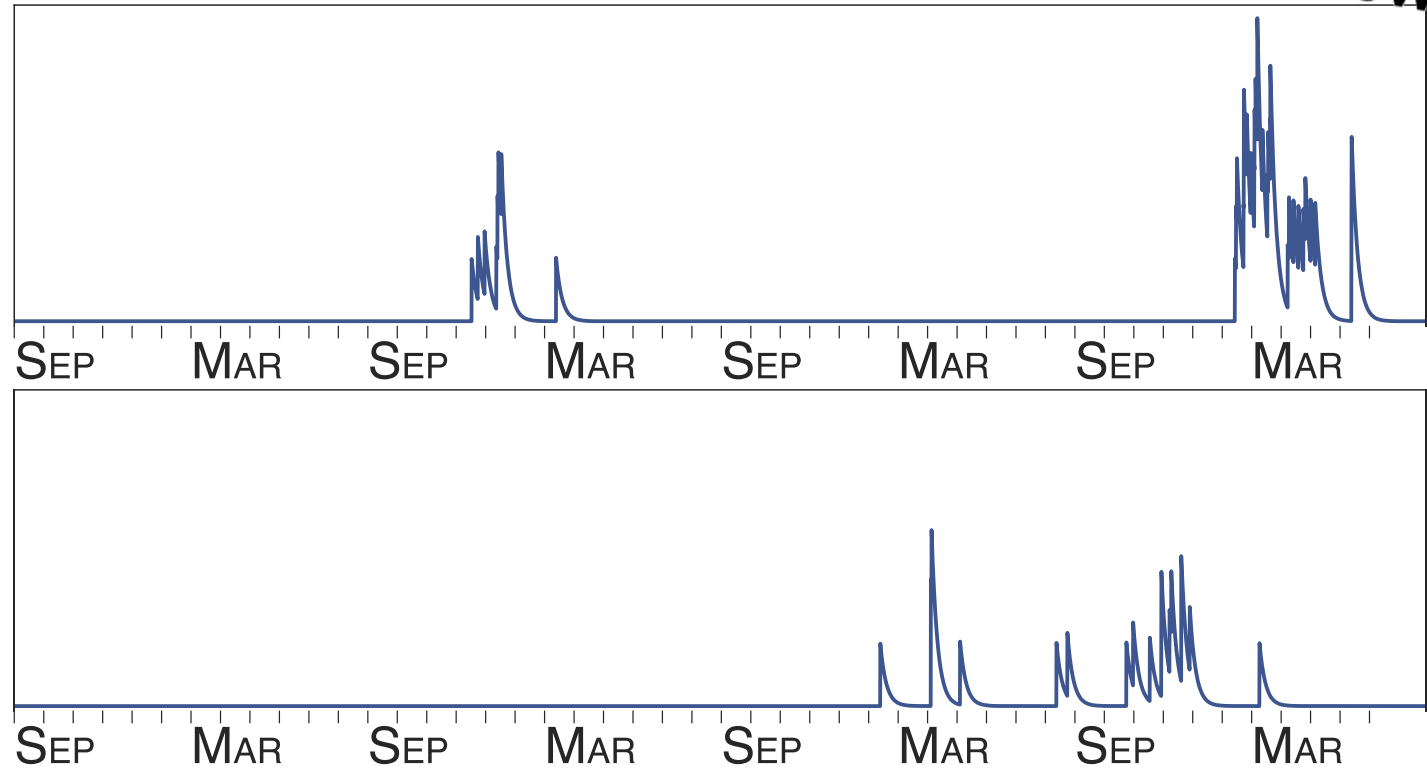
Machine learning tasks tend to be more complex and require asking more questions

[Mavroforakis et al., WWW 2017]

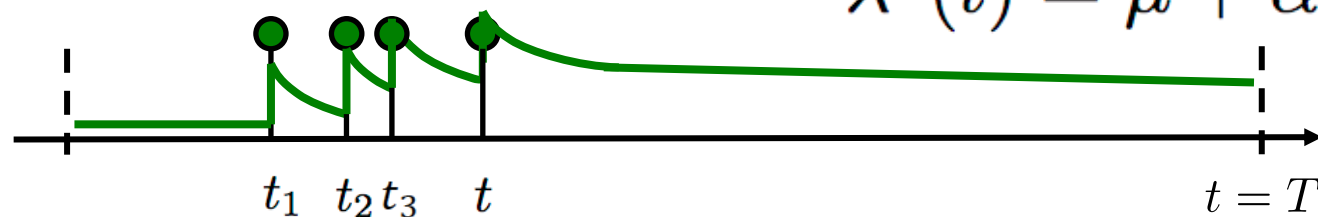# Models & Inference

Up to now, we have focused on simple temporal dynamics (and intensity functions):

$$\lambda^*(t) = \mu$$
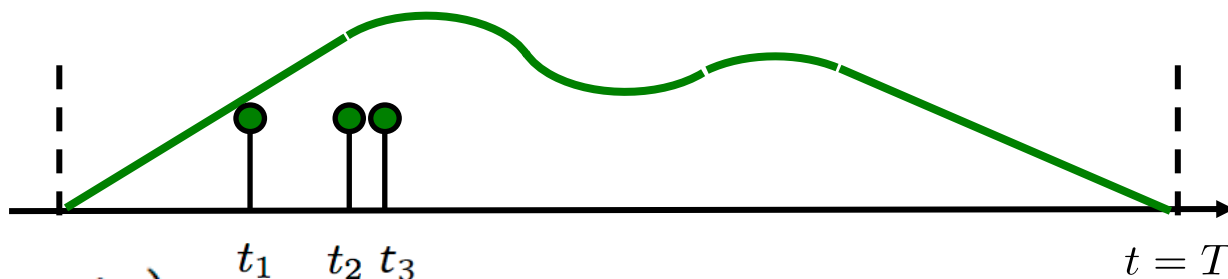
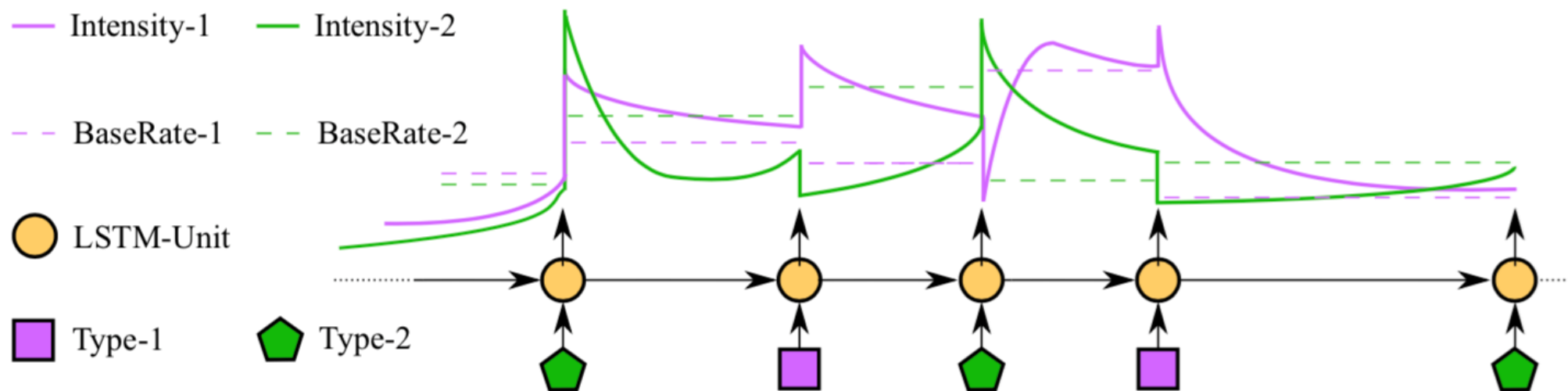$$\lambda^*(t) = \sum_j \alpha_j \, k(t - t_j)$$

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

Recent works make use of RNNs to capture more complex dynamics

[Du et al., 2016; Dai et al., 2016; Mei & Eisner, 2017; Jing & Smola, 2017; Trivedi et al., 2017; Xiao et al., 2017a; 2018]

Neural Hawkes process

1) History effect does not need to be additive
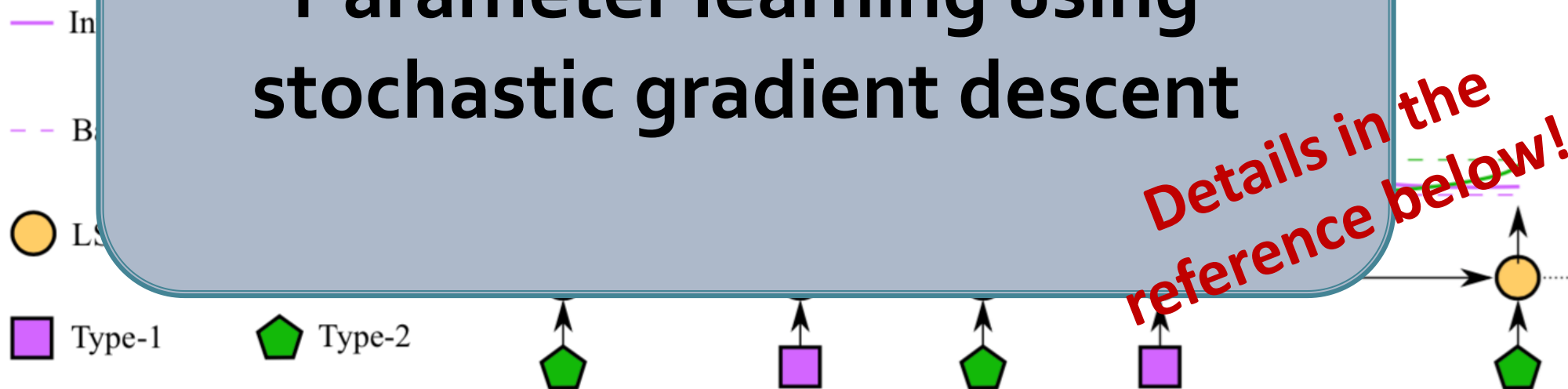
2) Allows for complex memory effects
   (such as delays)

[Mei & Eisner, NIPS 2017]

Neural Hawkes process

$$\lambda_u(t) = f_u(\mathbf{w}_u^\top \mathbf{h}(t))$$

Memory

$$\mathbf{h}(t) = \mathrm{RNN}(\mathcal{H}(t))$$
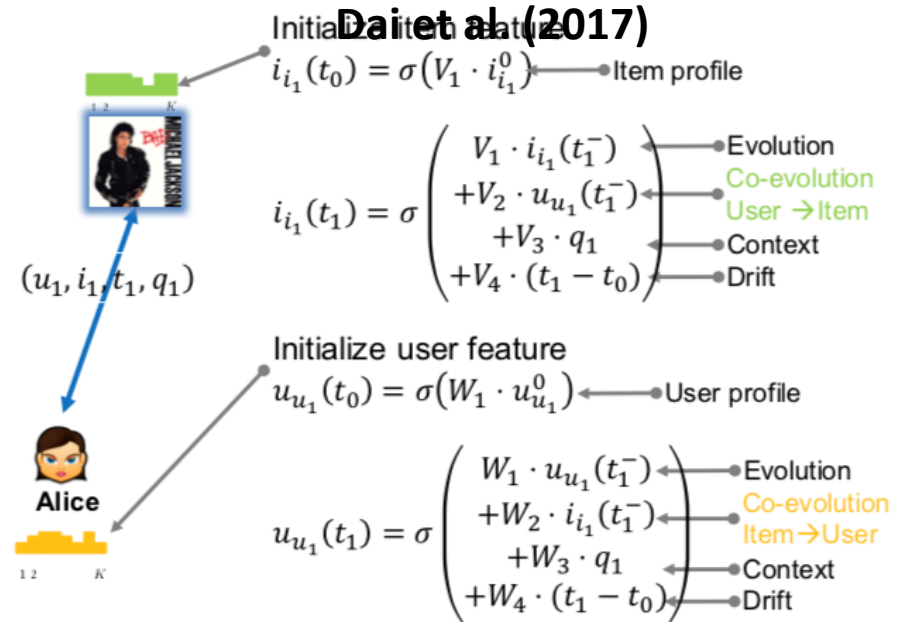
Parameter learning using stochastic gradient descent

Details in the reference below!

In

B

LS

■ Type-1   ⬟ Type-2

[Mei & Eisner, NIPS 2017]

Applications (I): Predictive Models

Know-Evolve, Trivedi et al. (2017)
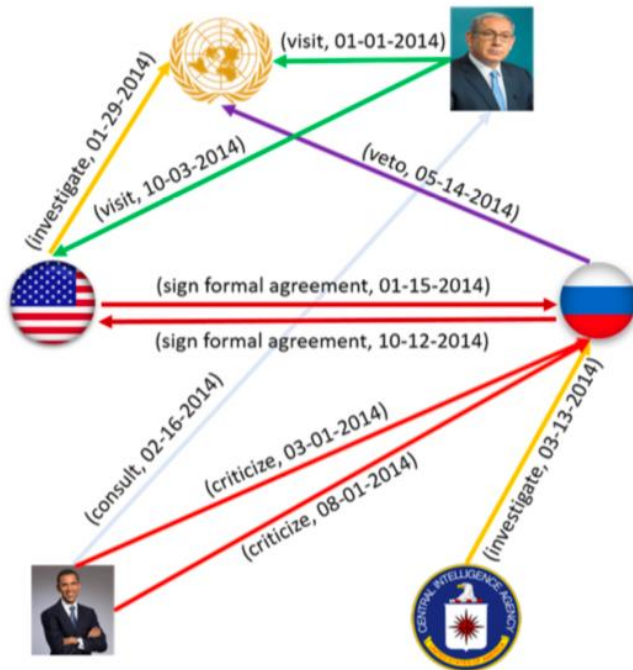
Coevolutionary Embedding, Dai et al. (2017)

Initialize item feature
$$i_{i_1}(t_0) = \sigma(V_1 \cdot i_{i_1}^0)$$ — Item profile

$$i_{i_1}(t_1) = \sigma \begin{pmatrix} V_1 \cdot i_{i_1}(t_1^-) \\ +V_2 \cdot u_{u_1}(t_1^-) \\ +V_3 \cdot q_1 \\ +V_4 \cdot (t_1 - t_0) \end{pmatrix} \begin{matrix} \text{Evolution} \\ \text{Co-evolution} \\ \text{User} \to \text{Item} \\ \text{Context} \\ \text{Drift} \end{matrix}$$

(u_1, i_1, t_1, q_1)

Alice

Initialize user feature
$$u_{u_1}(t_0) = \sigma(W_1 \cdot u_{u_1}^0)$$ — User profile

$$u_{u_1}(t_1) = \sigma \begin{pmatrix} W_1 \cdot u_{u_1}(t_1^-) \\ +W_2 \cdot i_{i_1}(t_1^-) \\ +W_3 \cdot q_1 \\ +W_4 \cdot (t_1 - t_0) \end{pmatrix} \begin{matrix} \text{Evolution} \\ \text{Co-evolution} \\ \text{Item} \to \text{User} \\ \text{Context} \\ \text{Drift} \end{matrix}$$

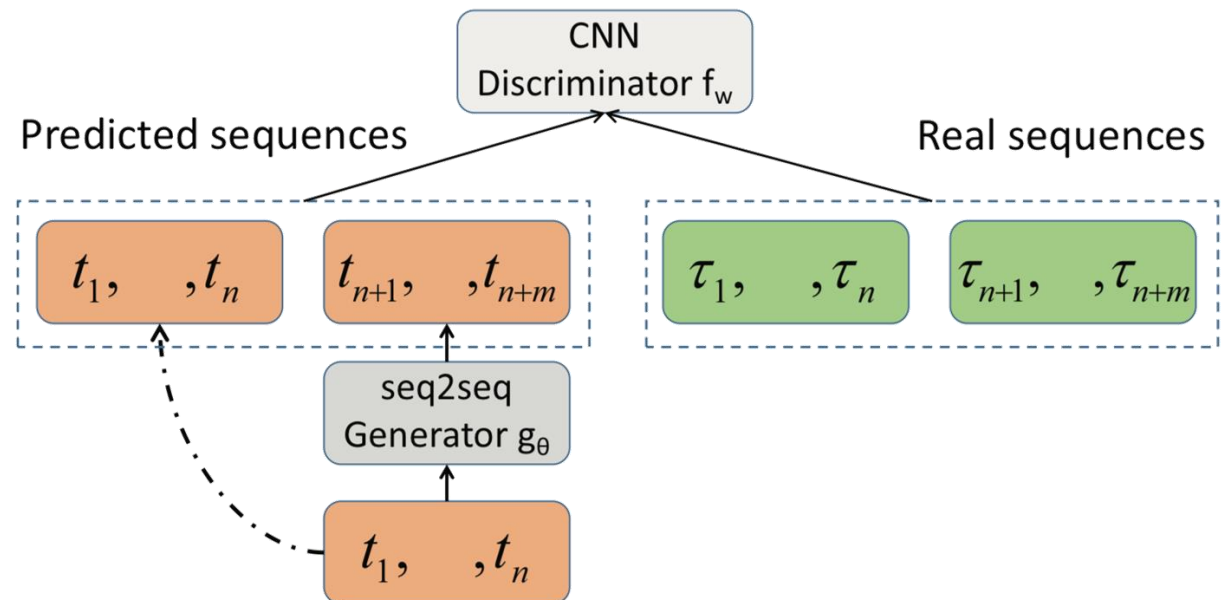**Neural Survival Recommender**, Jing & Smola (2017)
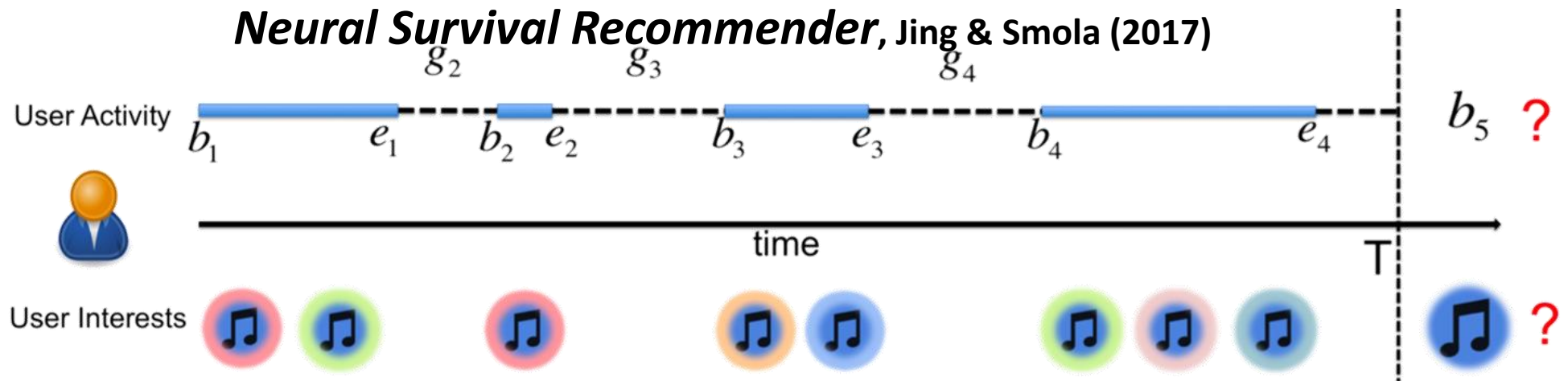
Key idea: Intensity- and likelihood-free models
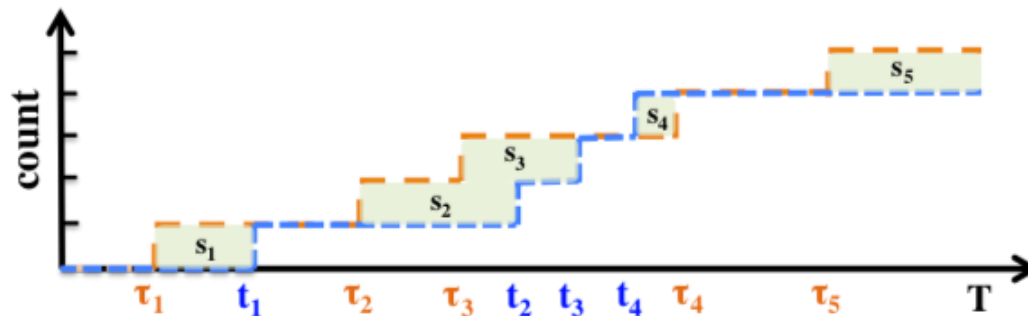
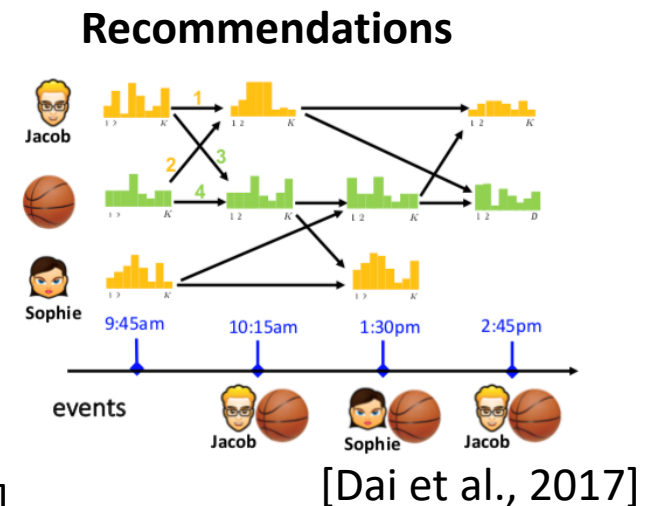**Wasserstein-Distance for Temporal Point Processes**
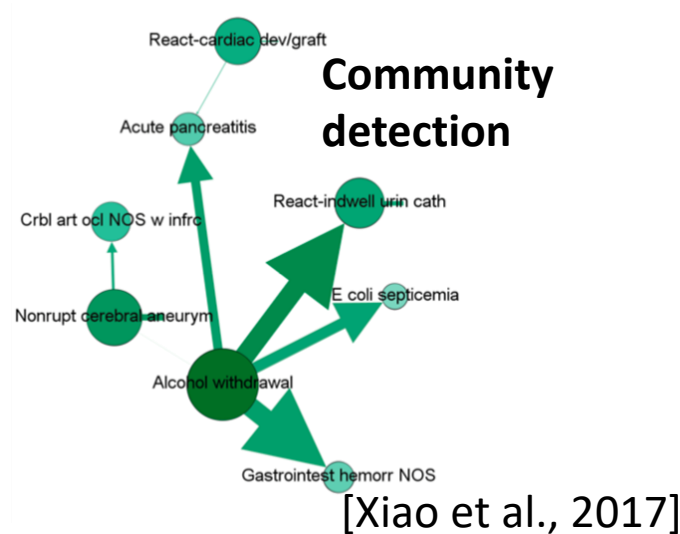
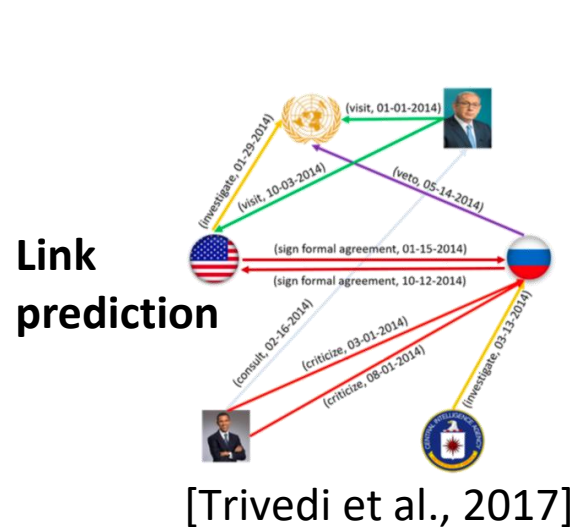**GAN architecture**

[Xiao et al., 2017 & 2018]
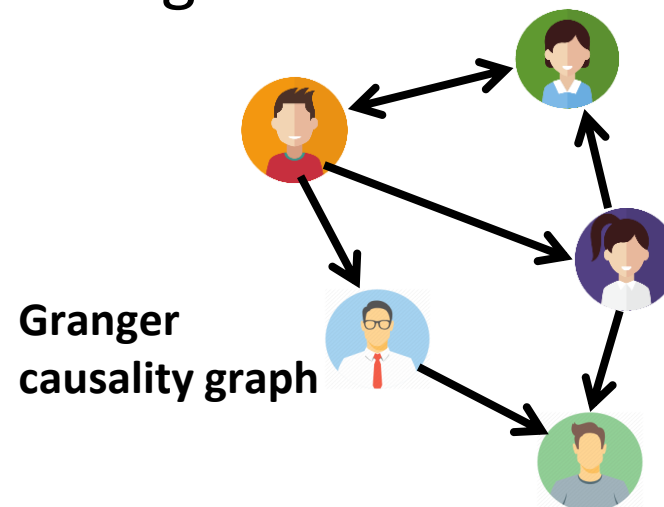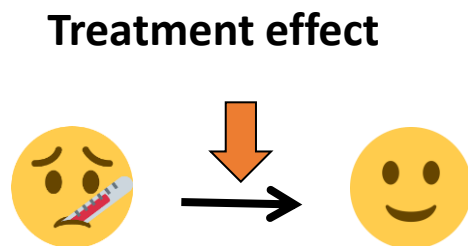
Models & Inference

1. Modeling event sequences
2. Clustering event sequences
3. Capturing complex dynamics
4. Causal reasoning on event sequences

Temporal point processes beyond prediction

So far, we have focused on models that improve preditions:

Link prediction

[Trivedi et al., 2017]

Community detection

React-cardiac dev/graft
Acute pancreatitis
React-indwell urin cath
Crbl art ocl NOS w infrc
E coli septicemia
Nonrupt cerebral aneurym
Alcohol withdrawal
Gastrointest hemorr NOS

[Xiao et al., 2017]

Recommendations

Jacob
Sophie
9:45am   10:15am   1:30pm   2:45pm
events
Jacob   Sophie   Jacob

[Dai et al., 2017]

Recent works have focused on performing **causal inference using event sequences**:

Treatment effect

Granger causality graph

[Xu et al., 2016; Achab et al., 2017; Kuśmierczyk & Gomez-Rodriguez, 2018]
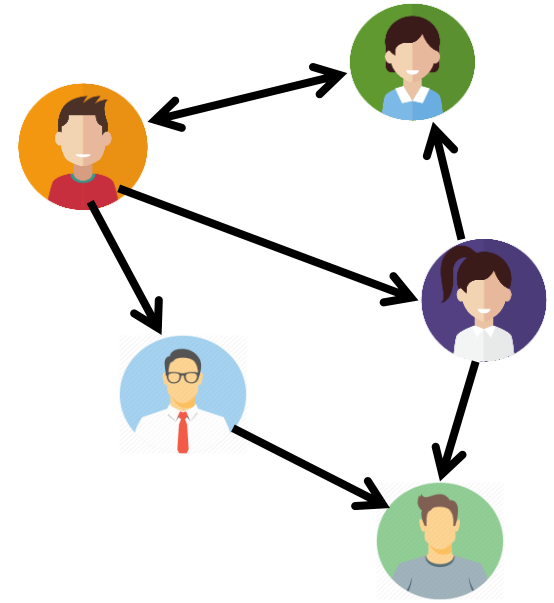
Uncovering Causality from Hawkes Processes

Multivariate Hawkes process:

$$N(t) = \sum_{u \in \mathcal{U}} N_u(t)$$

$$\lambda_u(t) = \mu_u + \sum_{v \in \mathcal{U}} \int_0^t \underbrace{k_{u,v}(t - t')dN_v(t')}_{\text{Effect of v's past events on u}}$$

Granger causality:

"X causes Y in the sense of Granger causality if forecasting future values of Y is more successful while taking X past values into account"
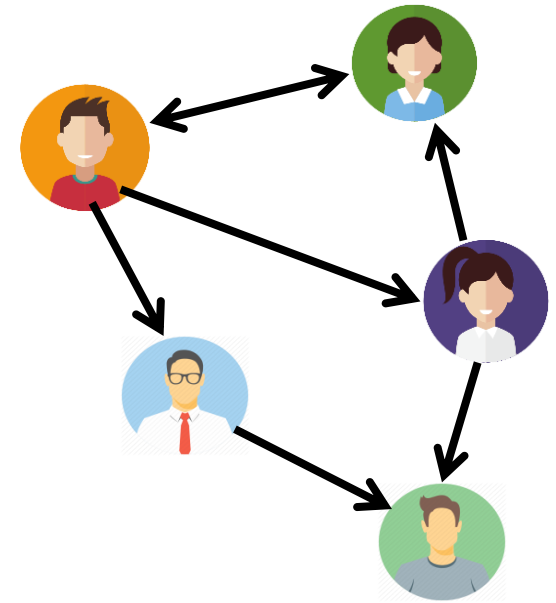
[Granger, 1969]

[Achab et al., ICML 2017]

Uncovering Causality from Hawkes Processes

Multivariate Hawkes process:

$$N(t) = \sum_{u \in \mathcal{U}} N_u(t)$$

$$\lambda_u(t) = \mu_u + \sum_{v \in \mathcal{U}} \int_0^t \underbrace{k_{u,v}(t - t')dN_v(t')}$$

Effect of v's past events on u

Granger causality on multivariate Hawkes processes:

" $N_v(t)$ does not Ganger-cause $N_u(t)$ w.r.t. $N(t)$ if and only if $k_{u,v}(\tau) = 0$ for $\tau \in \Re^+$ "
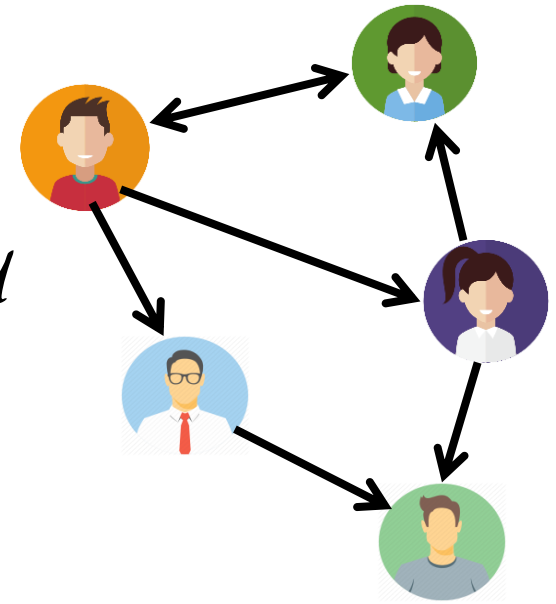
[Eichler et al., 2016]

[Achab et al., ICML 2017]

Uncovering Causality from Hawkes Processes

Goal is to estimate $G = [g_{uv}]$, where:

$$g_{uv} = \int_0^{+\infty} k_{u,v}(\tau)d\tau \geq 0 \text{ for all } u, v \in \mathcal{U}$$

Average total # of events of node u whose *direct* **ancestor is an event by node v**

Then, $G = [g_{uv}]$ quantifies the *direct causal relationship* **between nodes.**

[Achab et al., ICML 2017]

Uncovering Causality from Hawkes Processes

Goal is to estimate $G = [g_{uv}]$**, where:**

$$g_{uv} = \int_0^{+\infty} k_{u,v}(\tau)d\tau \geq 0 \text{ for all } u, v \in \mathcal{U}$$

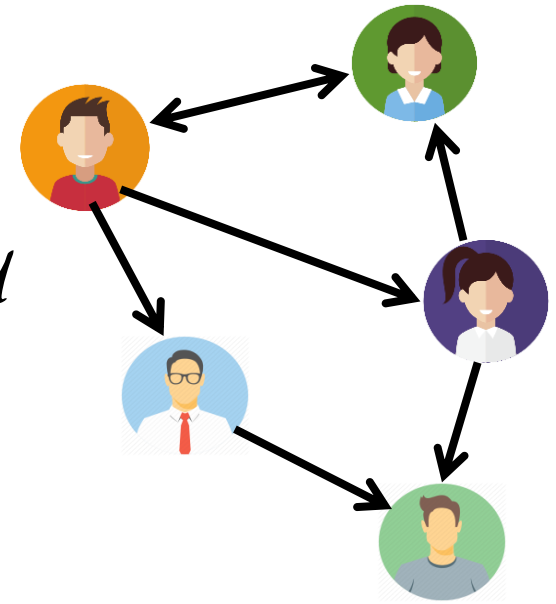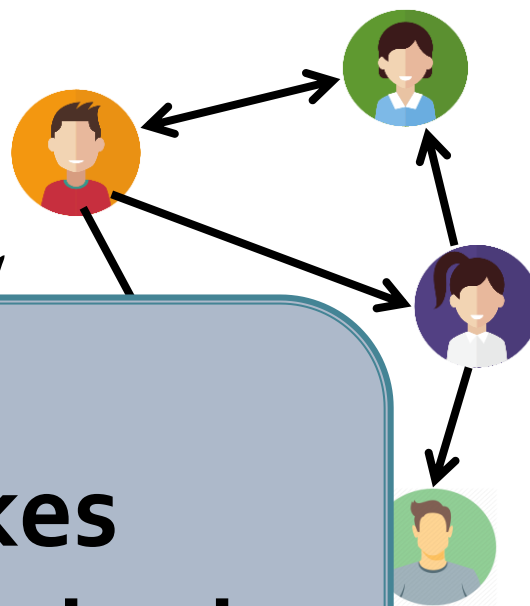**Average total # of events of node u whose** *direct* **ancestor is an event by node v**

Then, $G = [g_{uv}]$ **quantifies the** *direct causal relationship* **between nodes.**

Key idea: Estimate G using the cumulants dN(t) of the Hawkes process.

[Achab et al., ICML 2017]

Uncovering Causality from Hawkes Processes

Goal is to estimate $G = [g_{uv}]$**, where:**

$$g_{uv} = \int^{+\infty} k_{uv}(\tau)d\tau > 0 \text{ for all } u, v \in \mathcal{U}$$

Non parametric Hawkes cumulant estimation method
(with TensorFlow implementation)

Details in the hip reference below!

Key idea: **Estimate G using the cumulants the dN(t) of the Hawkes process.**

[Achab et al., ICML 2017]