Human-Centered ML
Part II

HUMAN-CENTERED MACHINE LEARNING

http://courses.mpi-sws.org/hcml-ws18/

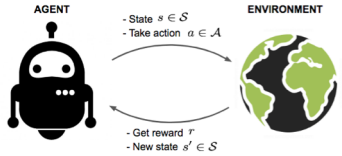
MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

Lectures for Part II

Five lectures on fundamentals

- Marked temporal Point Processes

- Optimal control
  Reinforcement learning



Five lectures on applications

Information propagation

Viral marketing

Opinion mining

Information integrity

duolingo

Human learning

Evaluation for Part II

Paper reviewing assignments:

- **Only for lectures on applications**
- **Due just before the lecture**

Two coding assignments:

- **Information propagation**
  From Dec 20 to Jan 17
- **Viral marketing**
  From Jan 24 to Feb 5

Final exam: **On Feb 7 (review on Feb 5)**

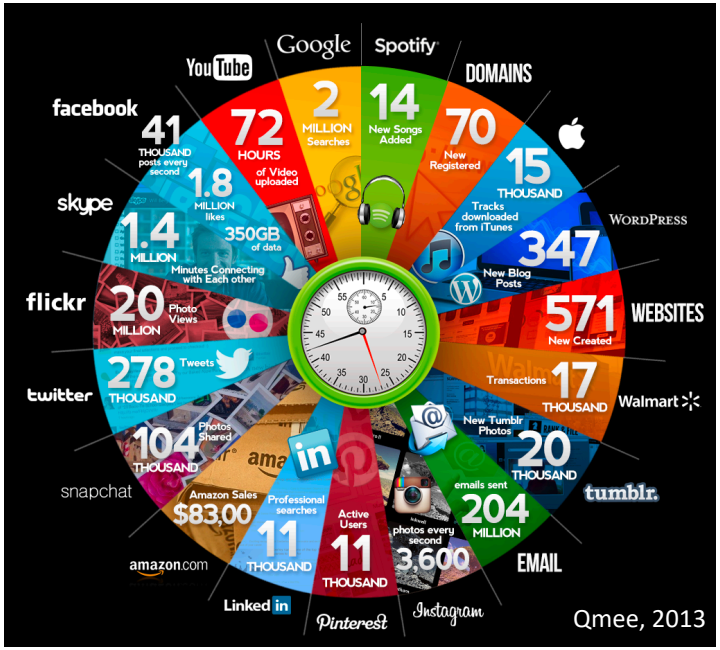
Introduction to
Temporal Point Processes (I)

HUMAN-CENTERED MACHINE LEARNING

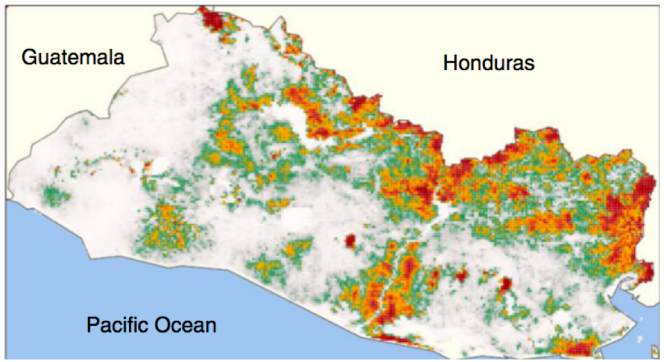
http://courses.mpi-sws.org/hcml-ws18/

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

Many discrete *events* in continuous time

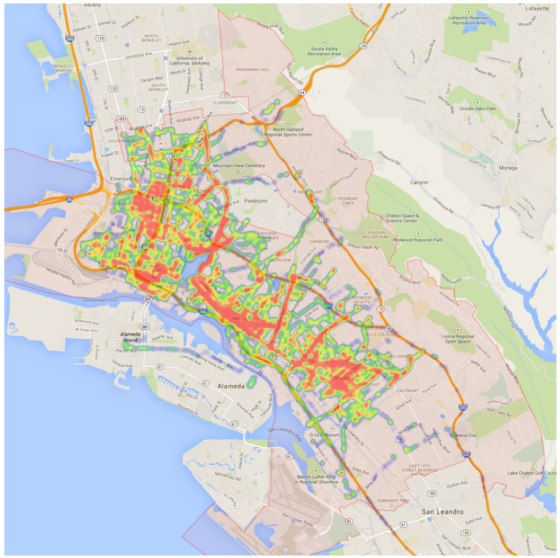


Online actions



Disease dynamics

Financial trading



Mobility dynamics

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

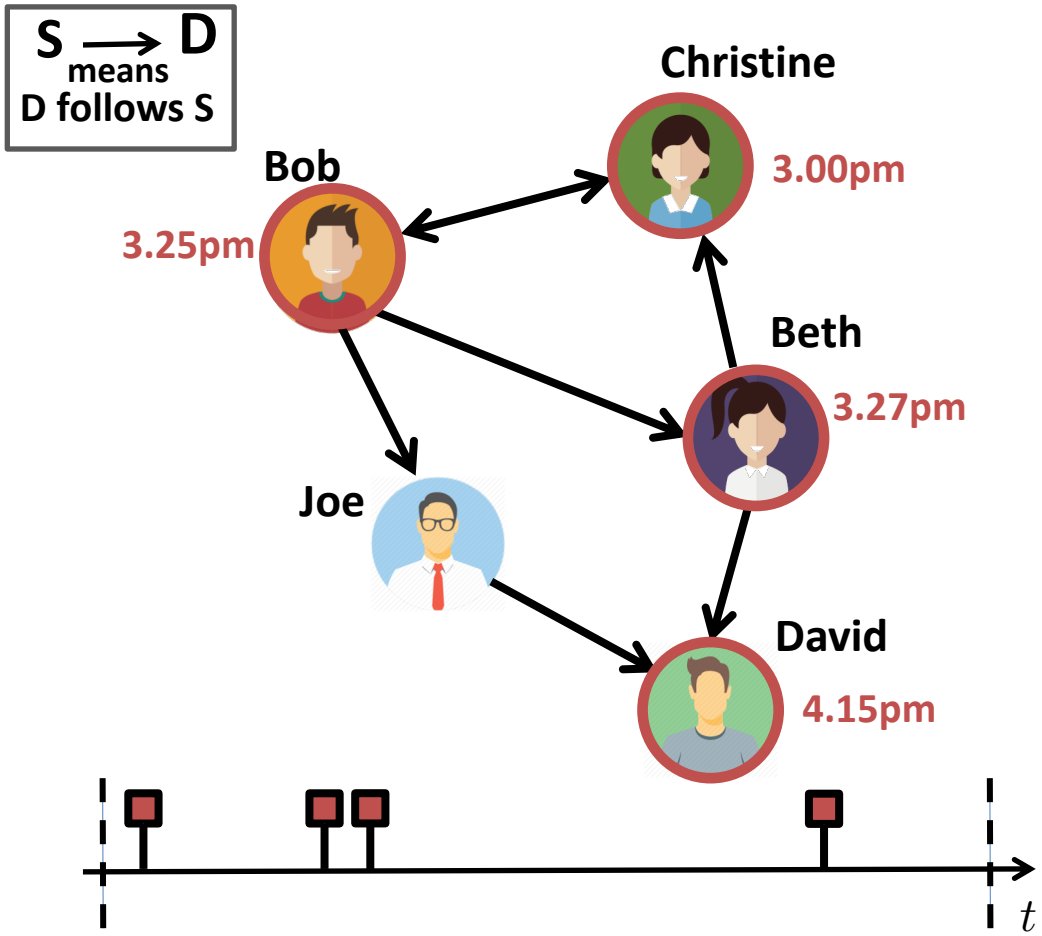
Ride-sharing requests

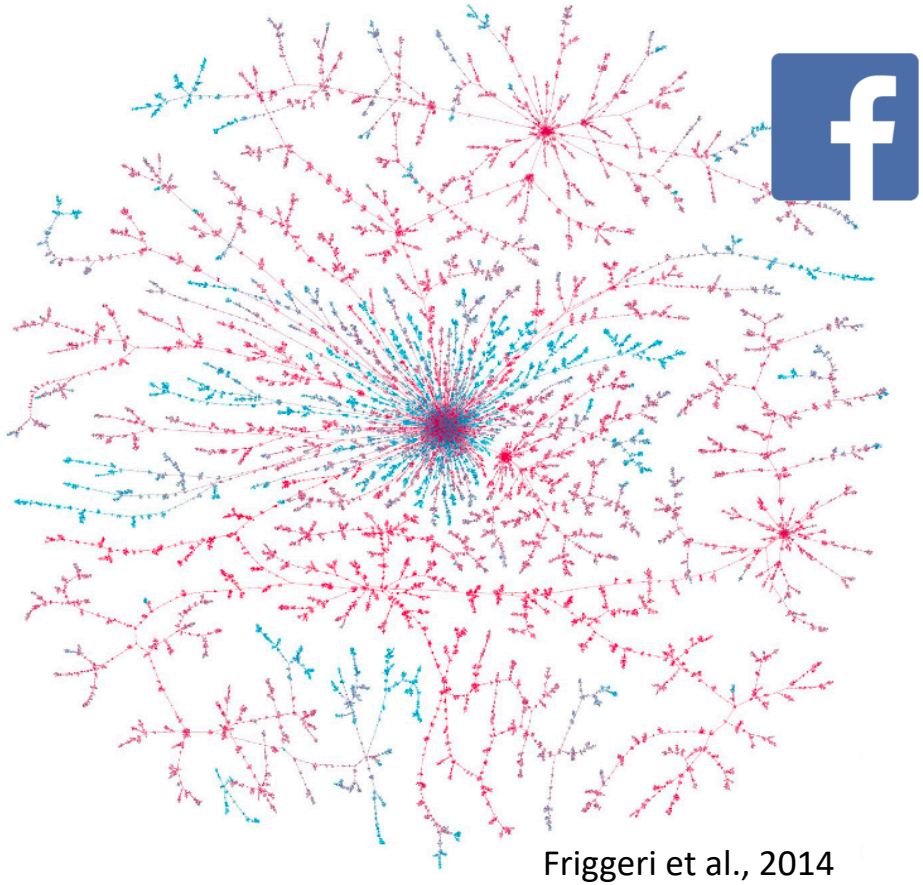
A user's reputation in Quora

FAST **SLOW**

…in a wide range of temporal scales.

Example I: Information propagation





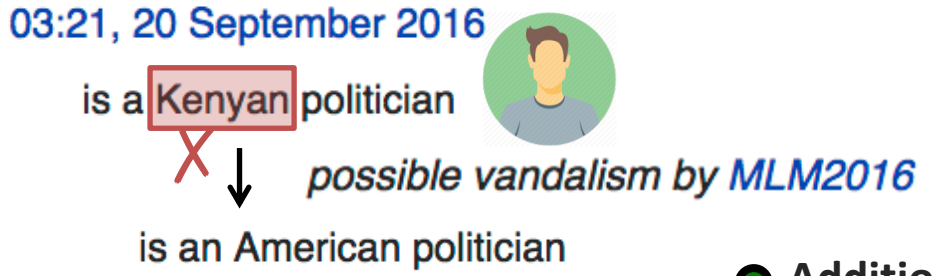
Friggeri et al., 2014

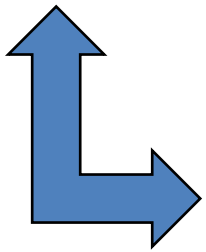
They can have an impact in the off-line world

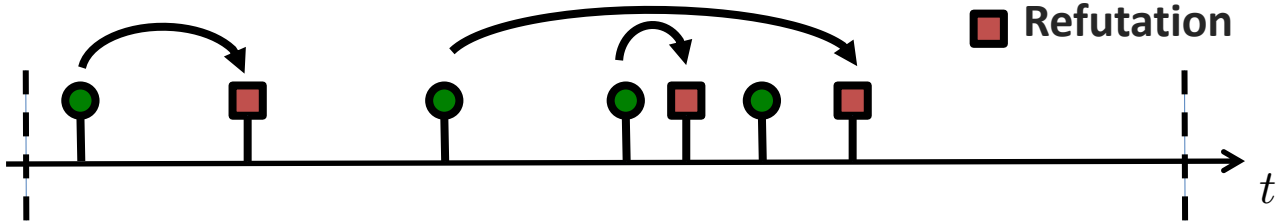
theguardian

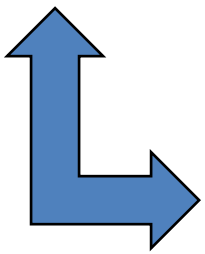
Click and elect: how fake news helped Donald Trump win a real election

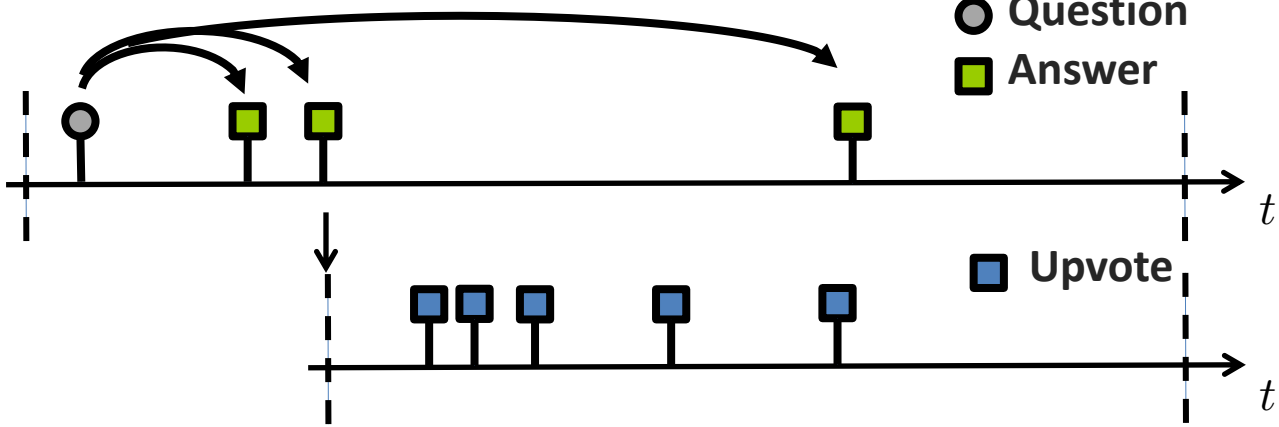
Example II: Knowledge creation





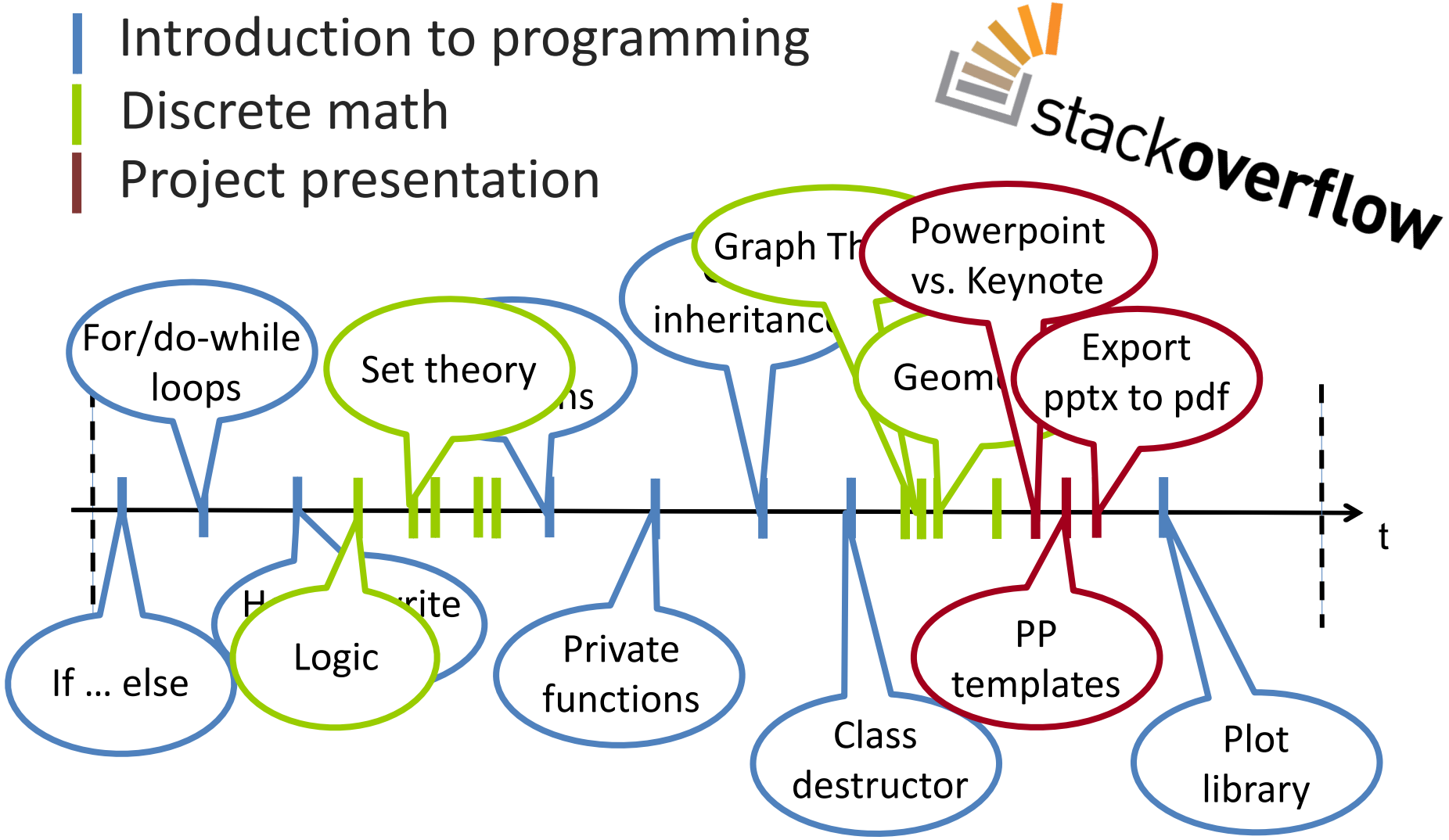






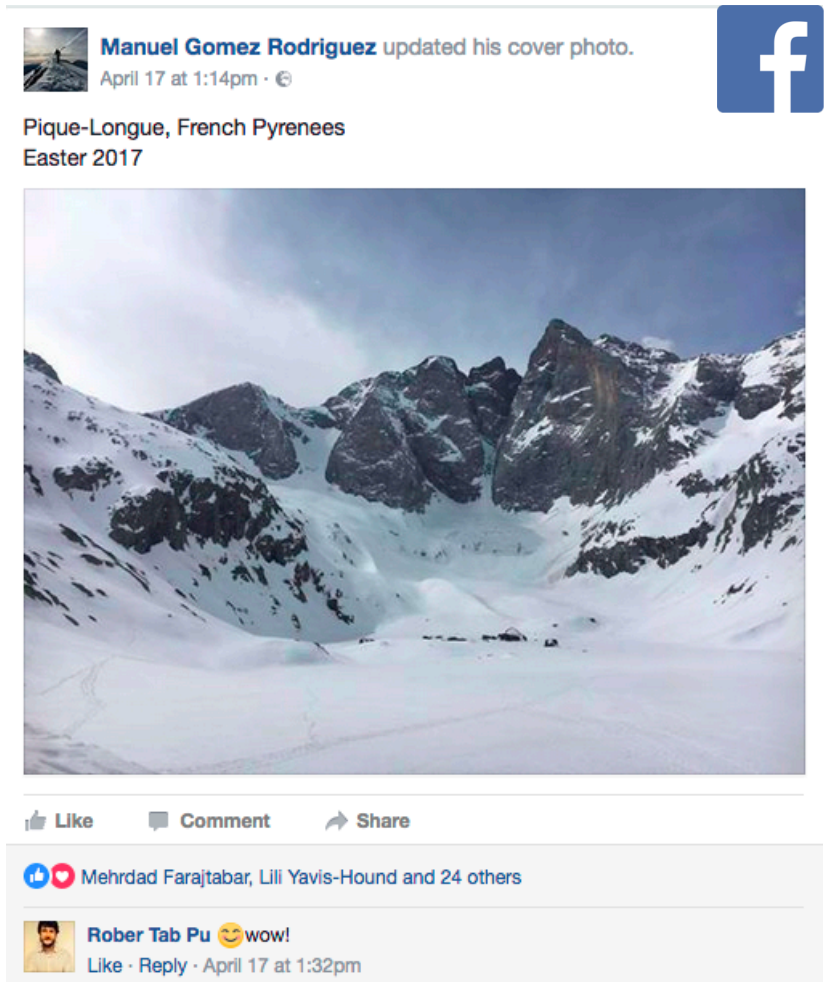
Example III: Human learning

1st year computer science student



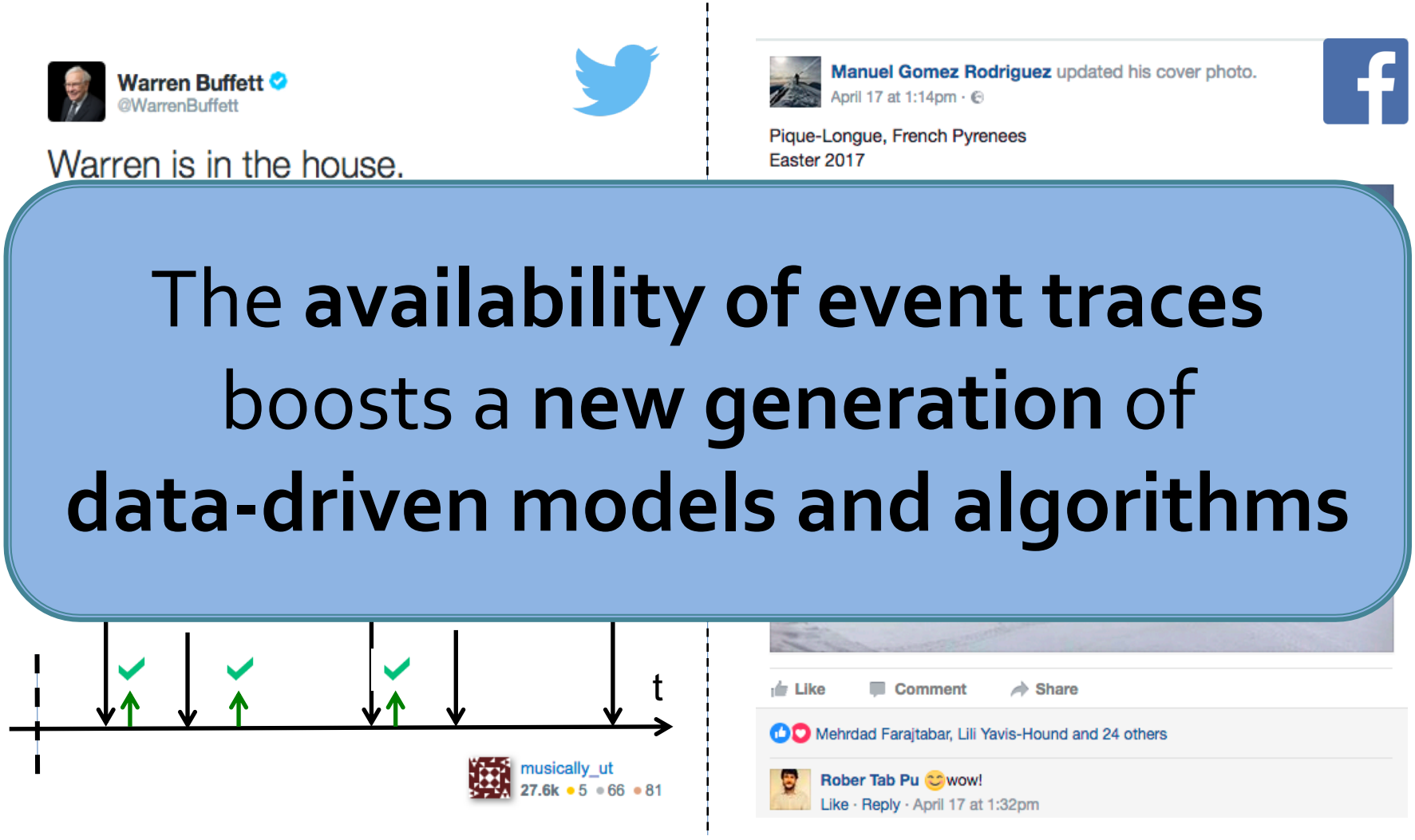
Detailed *event traces*

DETAILED TRACES OF ACTIVITY



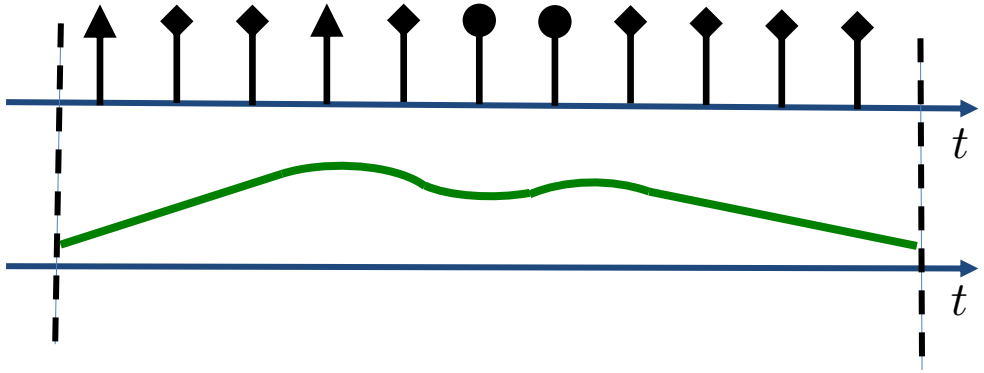
Detailed *event traces*

DETAILED TRACES OF ACTIVITY

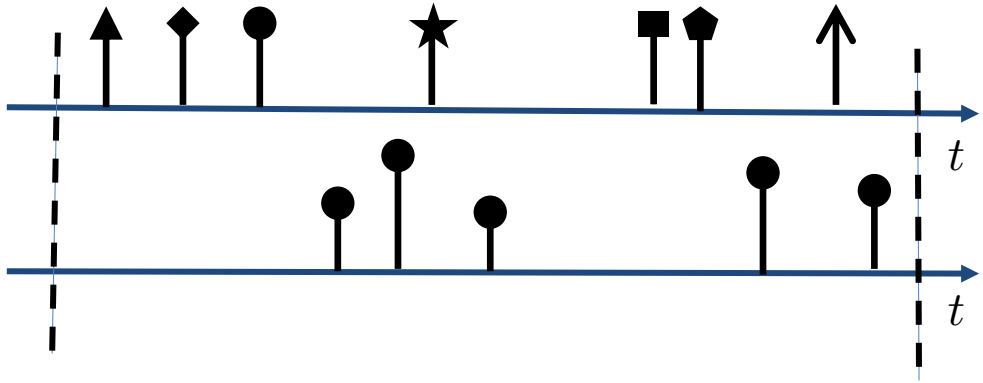


The **availability of event traces** boosts a **new generation** of **data-driven models and algorithms**

Aren't these event traces just time series?

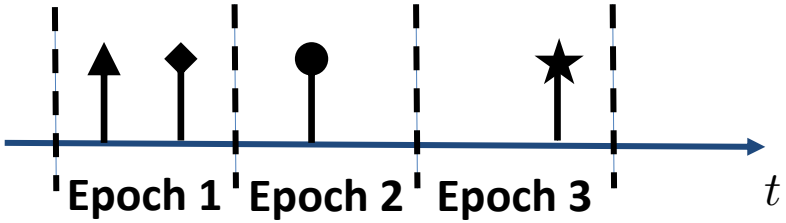


Discrete and continuous times series



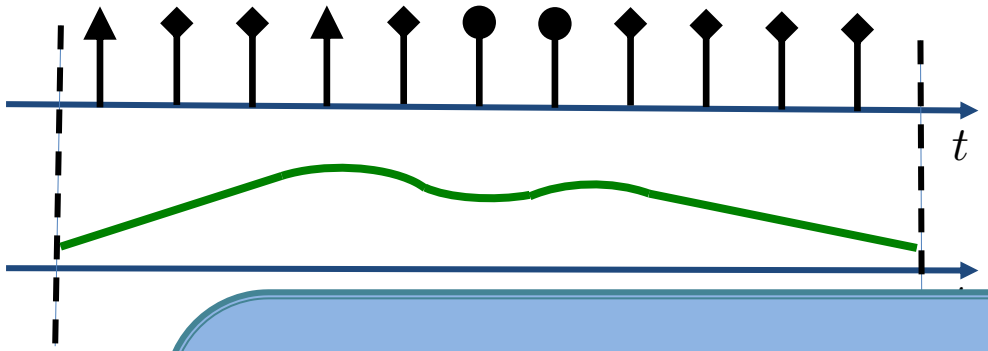
Discrete events in continuous time

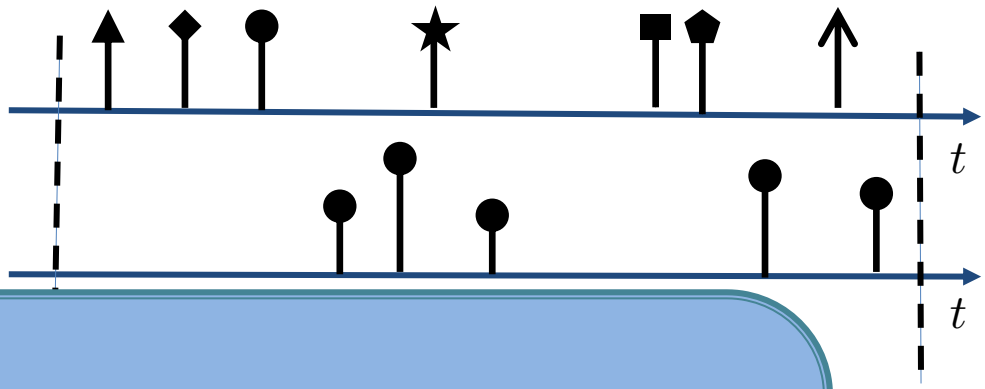
What about aggregating events in *epochs*?



- **How long is each epoch?**
- **How to aggregate events per epoch?**
- **What if no event in one epoch?**
- **What about time-related queries?**

Aren't these event traces just time series?





Dis

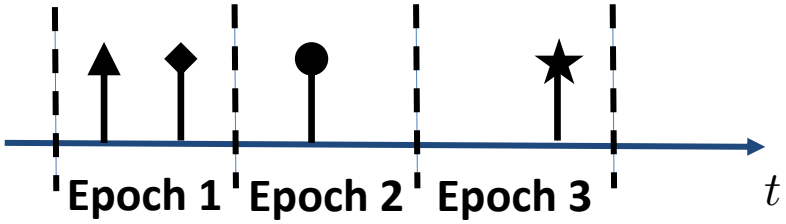
W

poch?

What if no event in one epoch?

What about time-related queries?

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

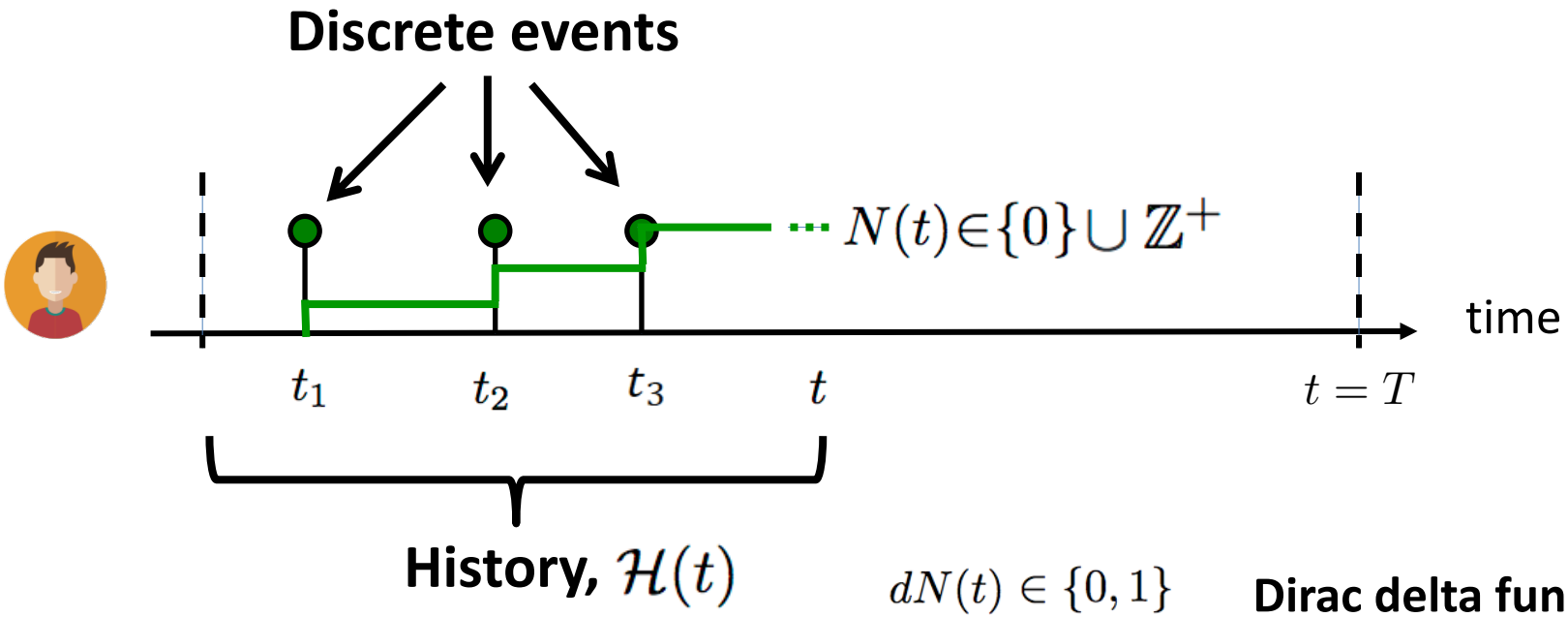


Temporal Point Processes:
Intensity function

Temporal point processes

Temporal point process:

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

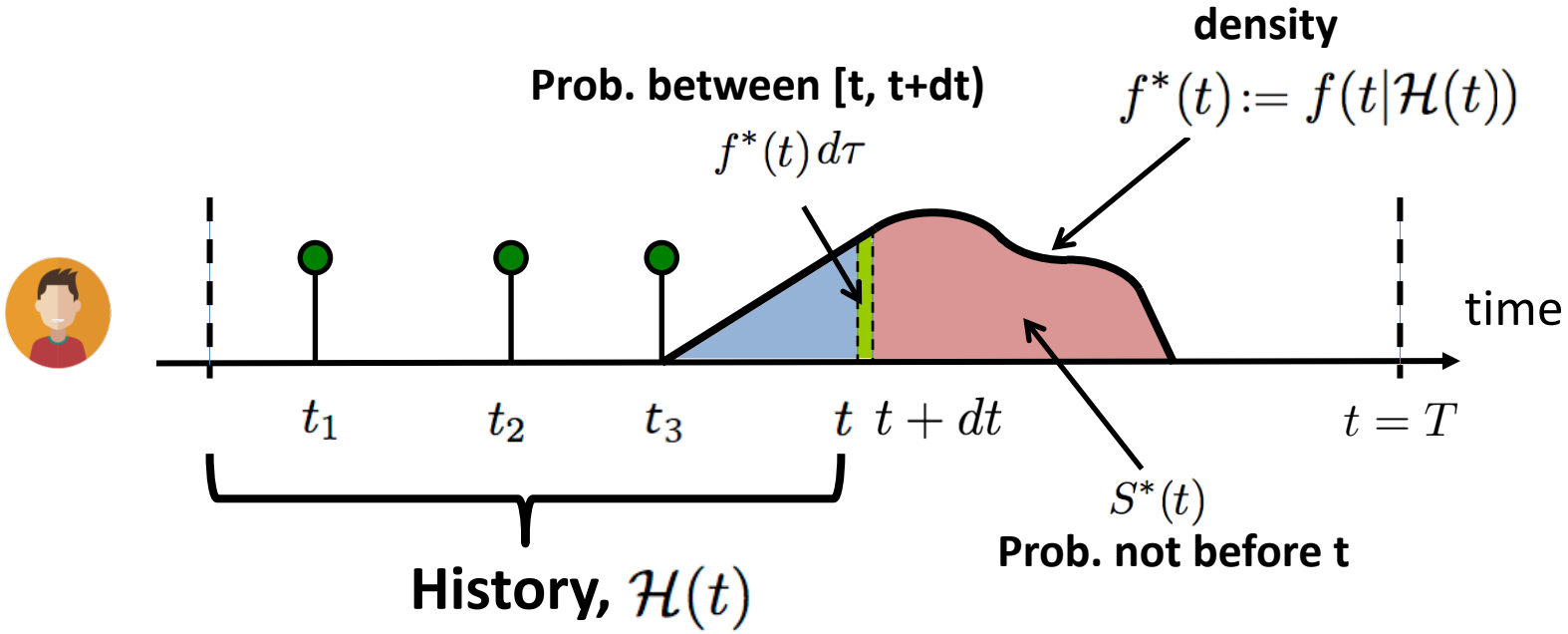


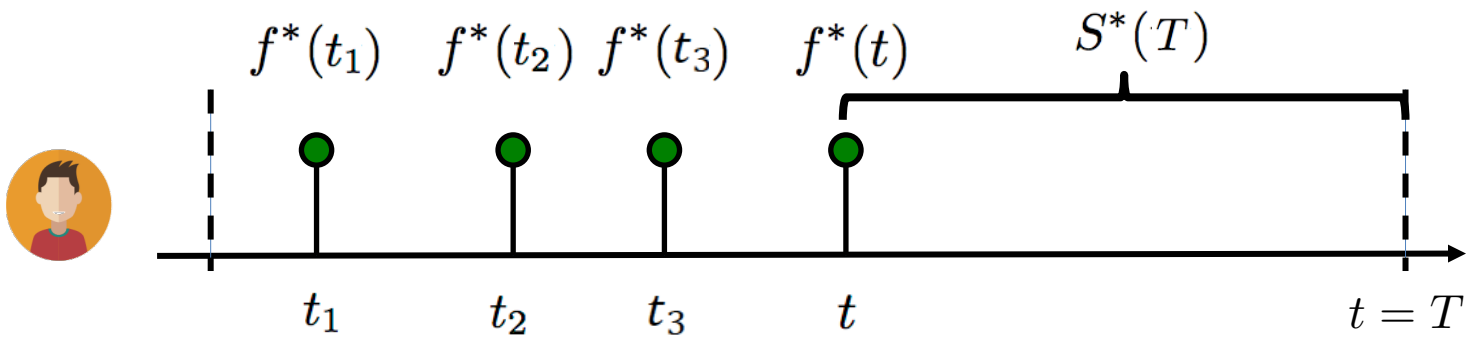
$dN(t) \in \{0, 1\}$

Dirac delta function

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

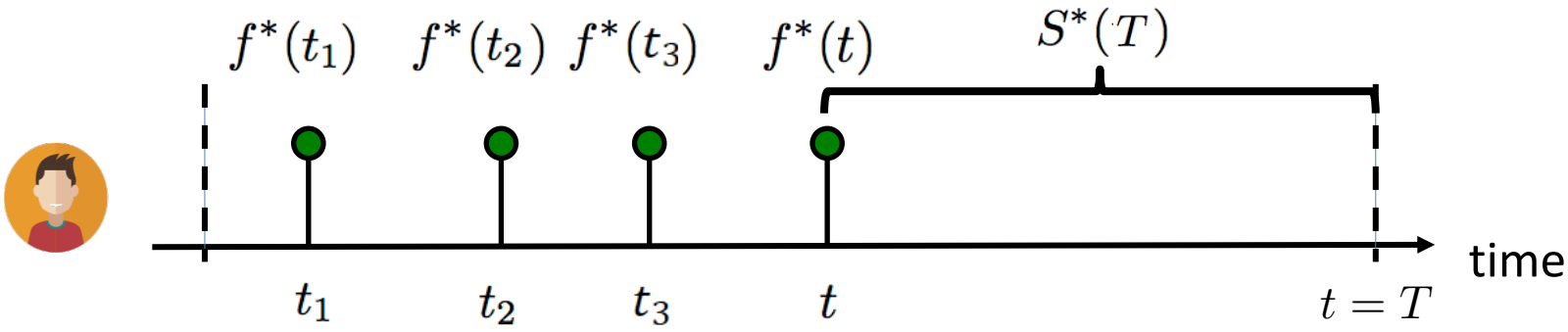
Model time as a random variable

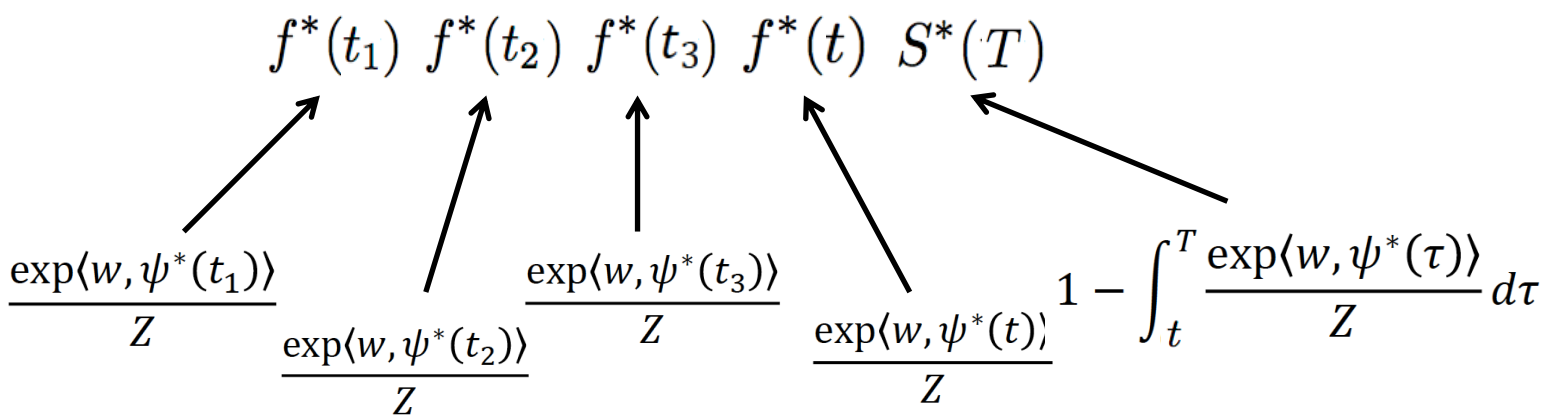




Likelihood of a timeline: $f^*(t_1)\ f^*(t_2)\ f^*(t_3)\ f^*(t)\ S^*(T)$

Problems of density parametrization (I)



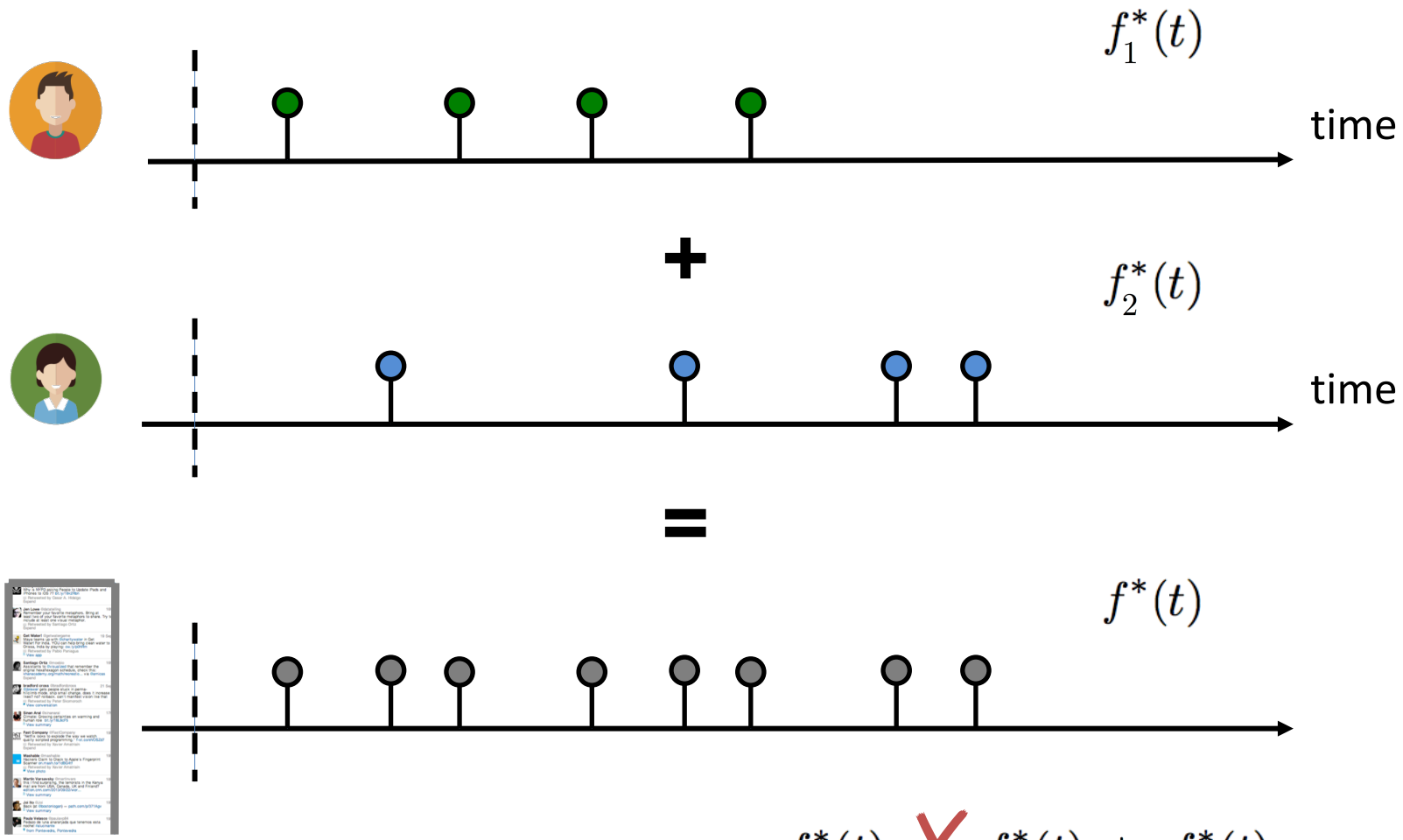


It is difficult for model design and interpretability:

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:

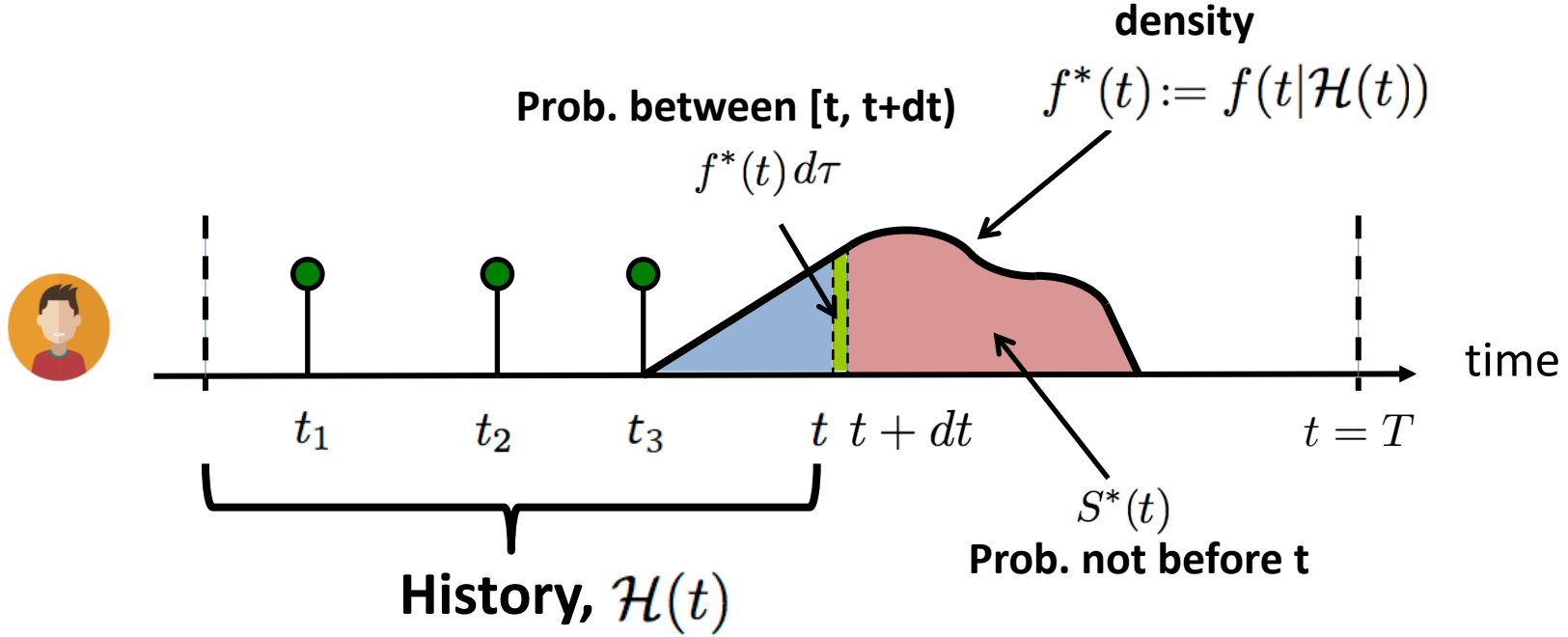


Sum of random processes

$$f^*(t) \neq f_1^*(t) + f_2^*(t)$$

$$f^*(t) \neq f_1^*(t) * f_2^*(t)$$

Intensity function



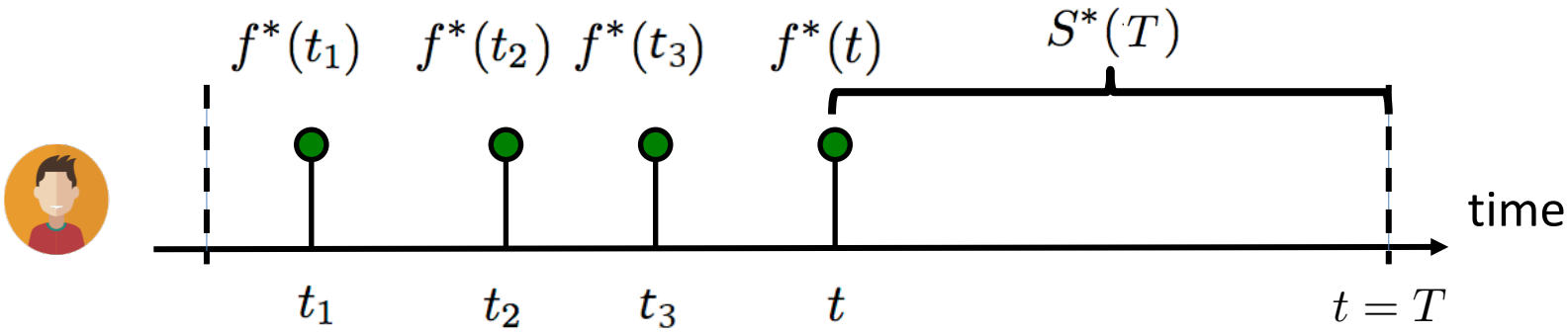
Intensity:

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)



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

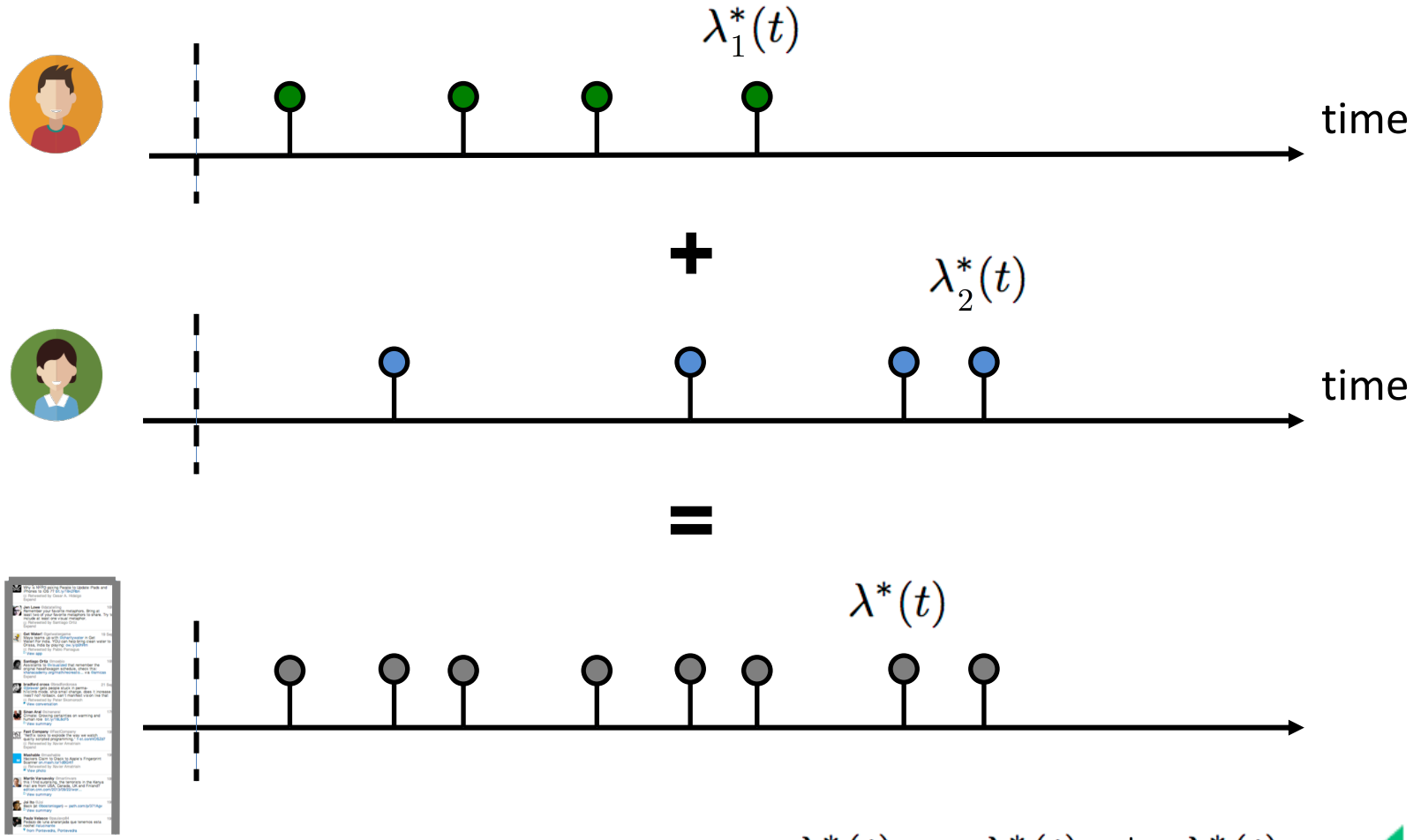
$\langle w, \phi^*(t_1)\rangle$ $\langle w, \phi^*(t_2)\rangle$ $\langle w, \phi^*(t_3)\rangle$ $\langle w, \phi^*(t)\rangle$ $\exp\left(-\int_0^T \langle w, \phi^*(\tau)\rangle d\tau\right)$

Suitable for model design and interpretable:

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

Advantages of intensity parametrization (II)

Easy to combine timeline:

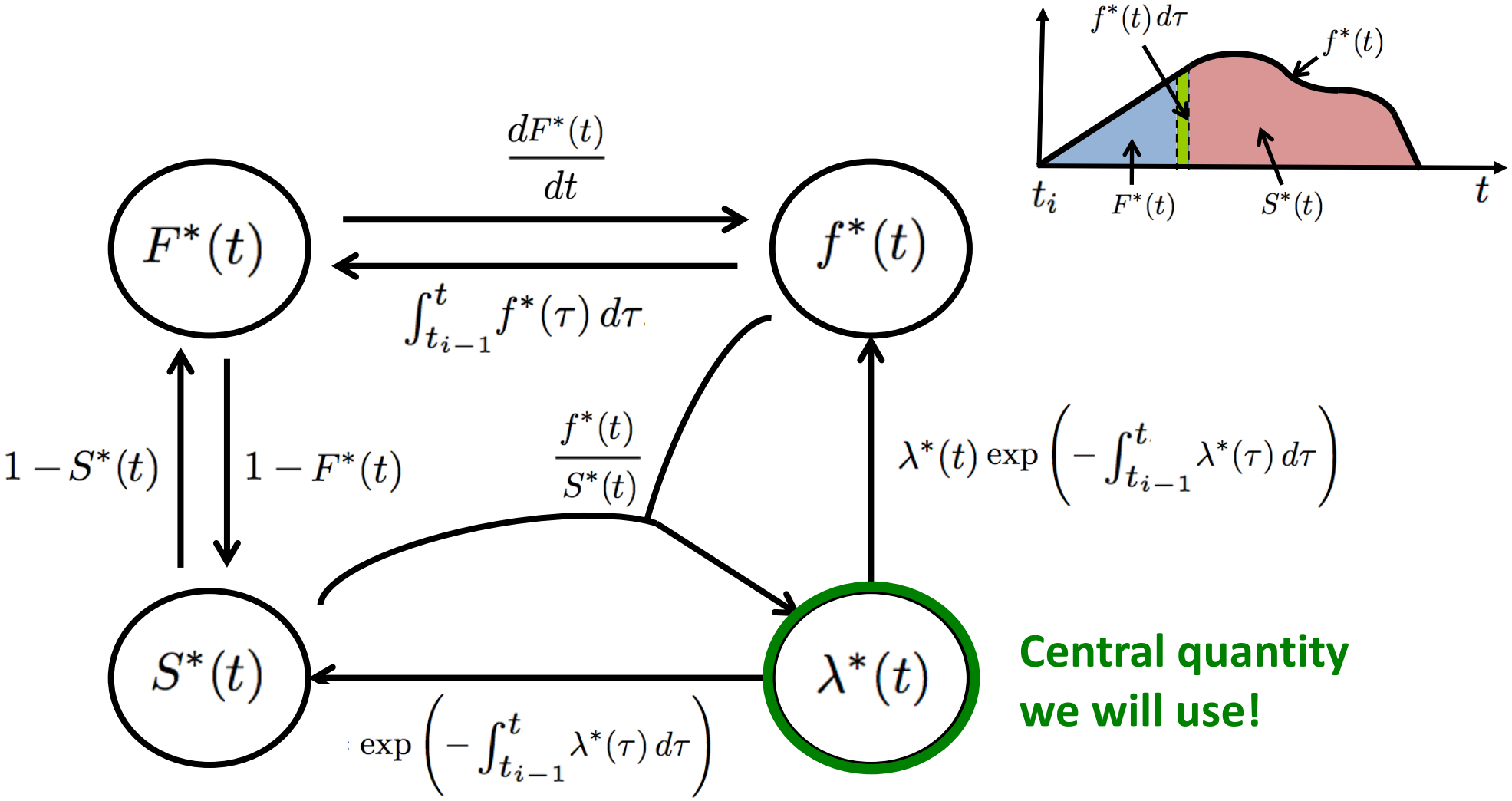


Sum of random processes

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

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

Relation between f*, F*, S*, λ*



Central quantity we will use!