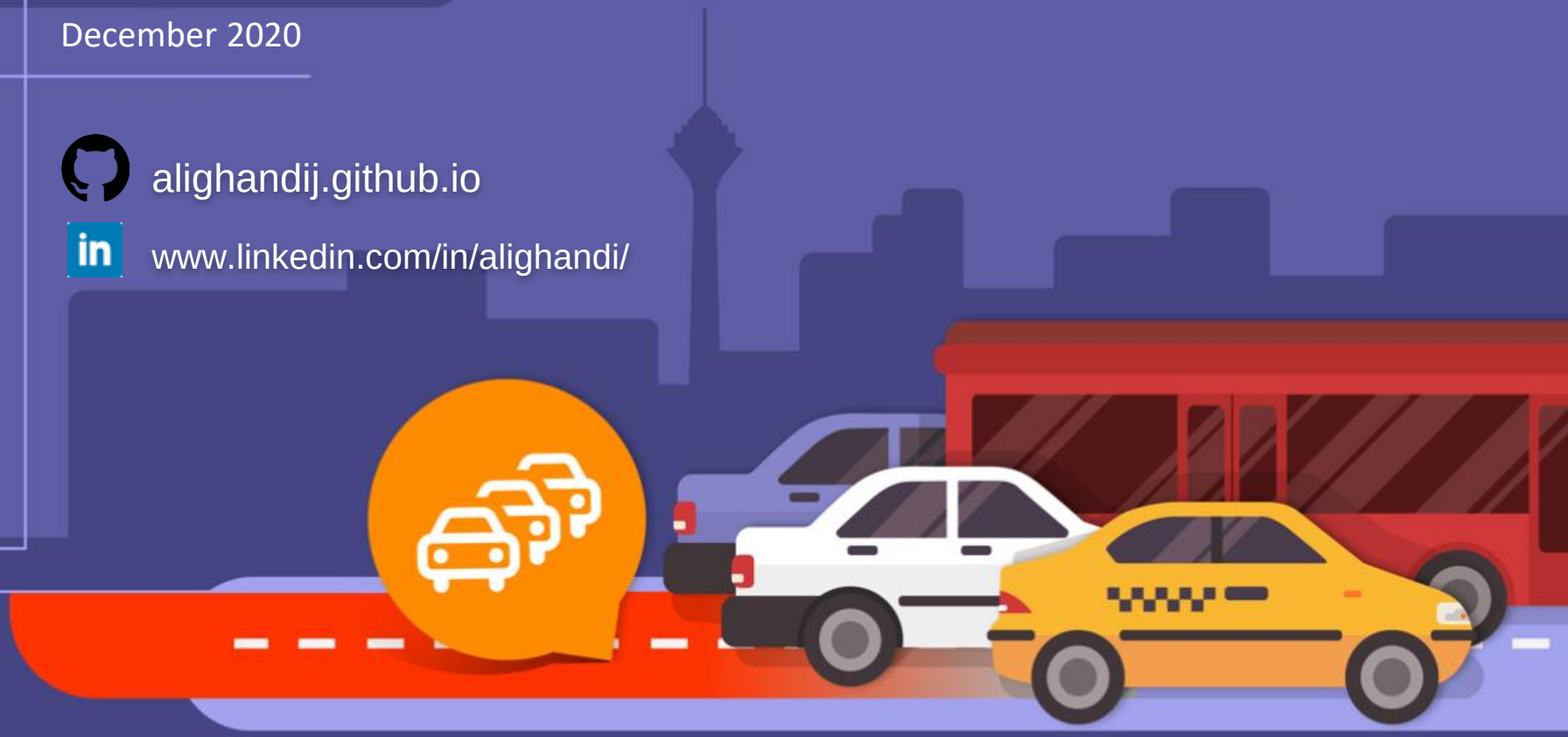
Live Traffic

Ali Ghandi

December 2020

alighandij.github.io

www.linkedin.com/in/alighandi/



گروه توسعه فناوری اطلاعات
هزاردستان

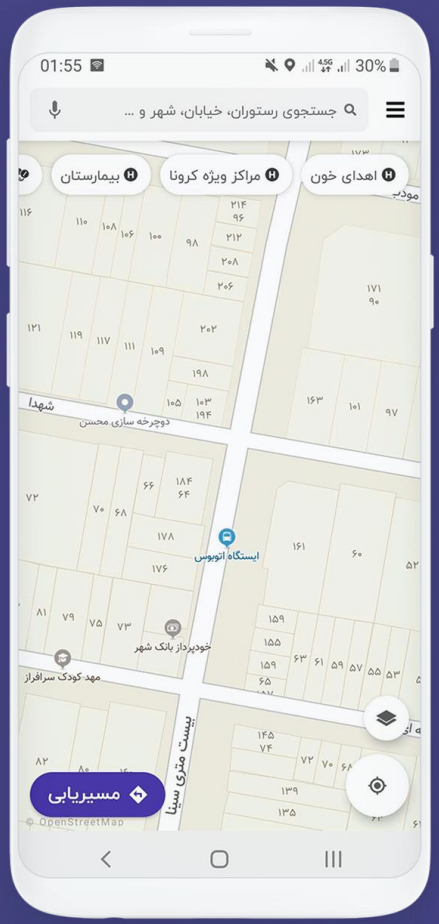
بازار

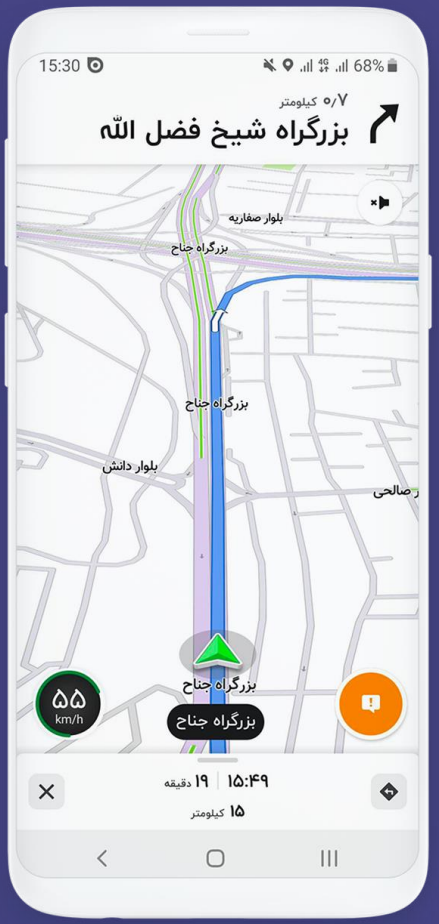
دیوار

بلد
نقشه و مسیریاب

Balad

- Start: **December 2016**
- Release: **May 2018**





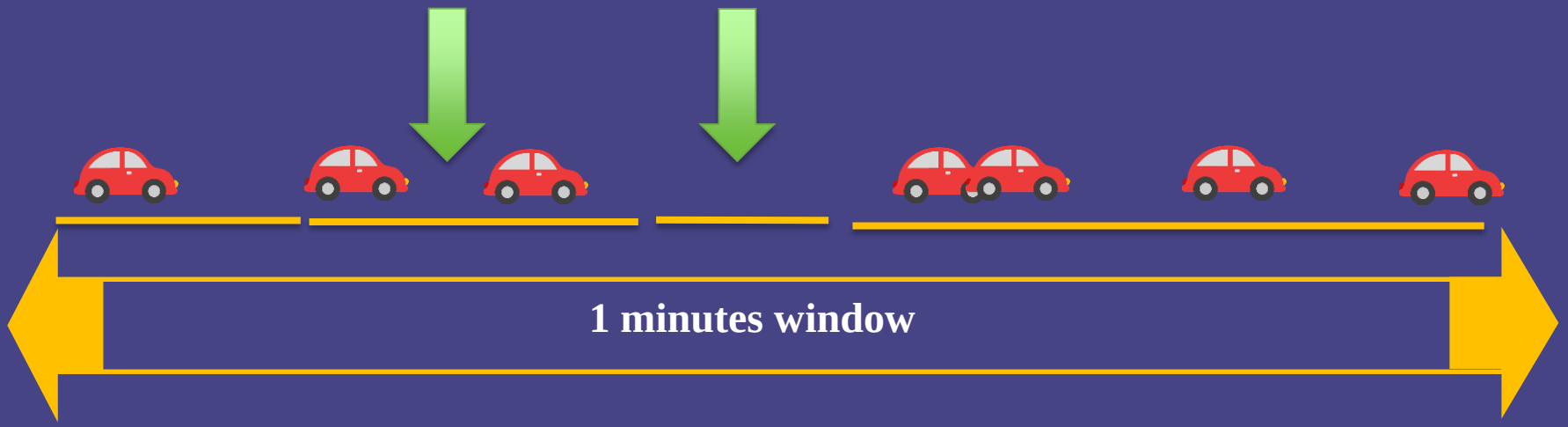
Balad

- Team members: **110**
- Users: **+5** million
- POI's: **+700,000**

What is Live?

$$\frac{\Delta x}{\Delta t}$$

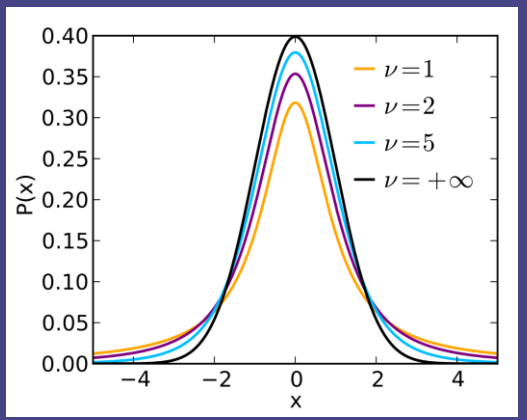
What is Live?



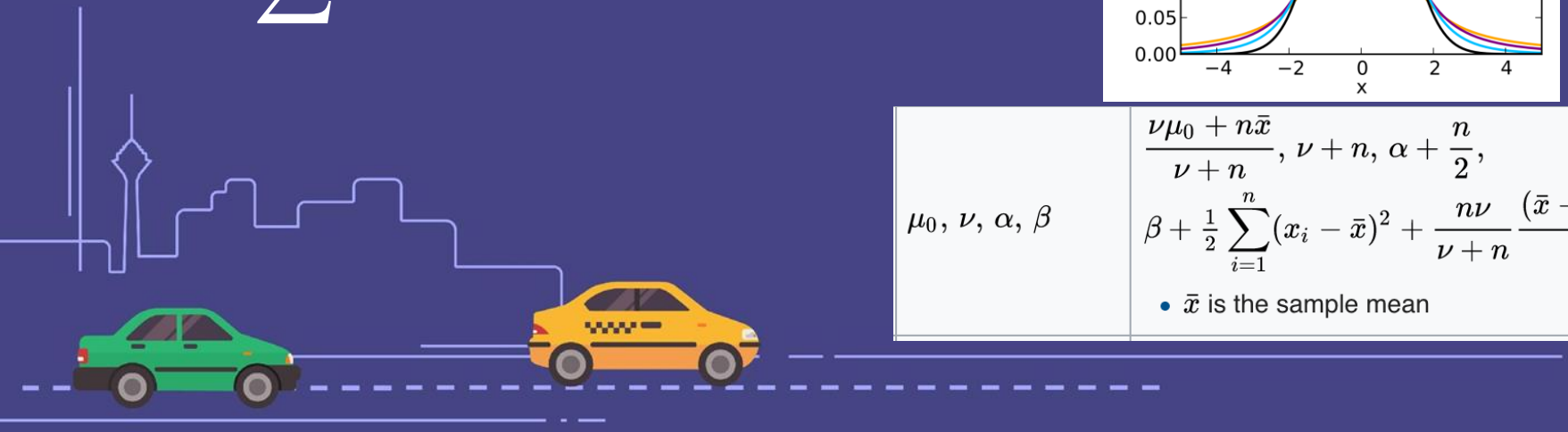
GPS location error

What is Live?

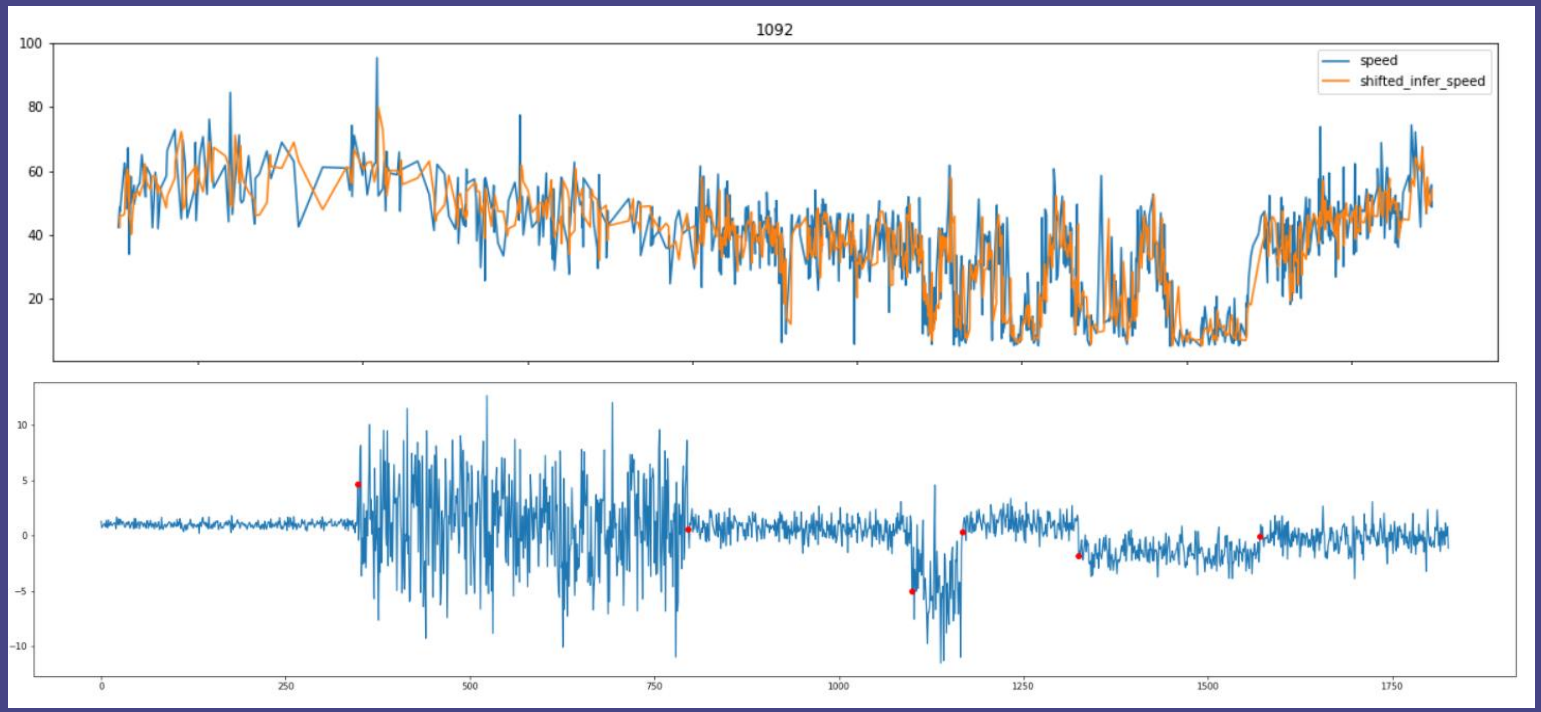
$$\sum \frac{e^{-\lambda t}{}_{\beta} v_t}{\sum e^{-\lambda t}} = V_{edge}$$



μ_0, ν, α, β	$\frac{\nu\mu_0 + n\bar{x}}{\nu + n}$, $\nu + n$, $\alpha + \frac{n}{2}$, $\beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{n\nu}{\nu + n}\frac{(\bar{x} - \mu_0)^2}{2}$ • $\bar{x}$ is the sample mean



Change point

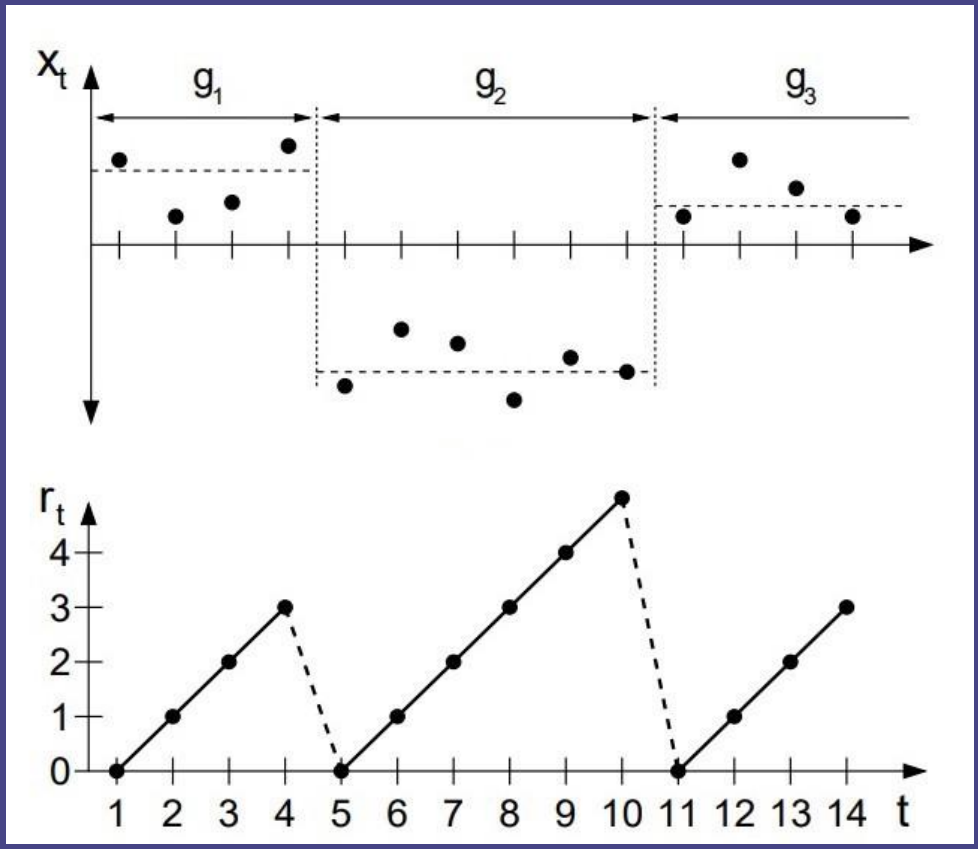


BOCD

Bayesian online change point detection

$$r_m(n)$$

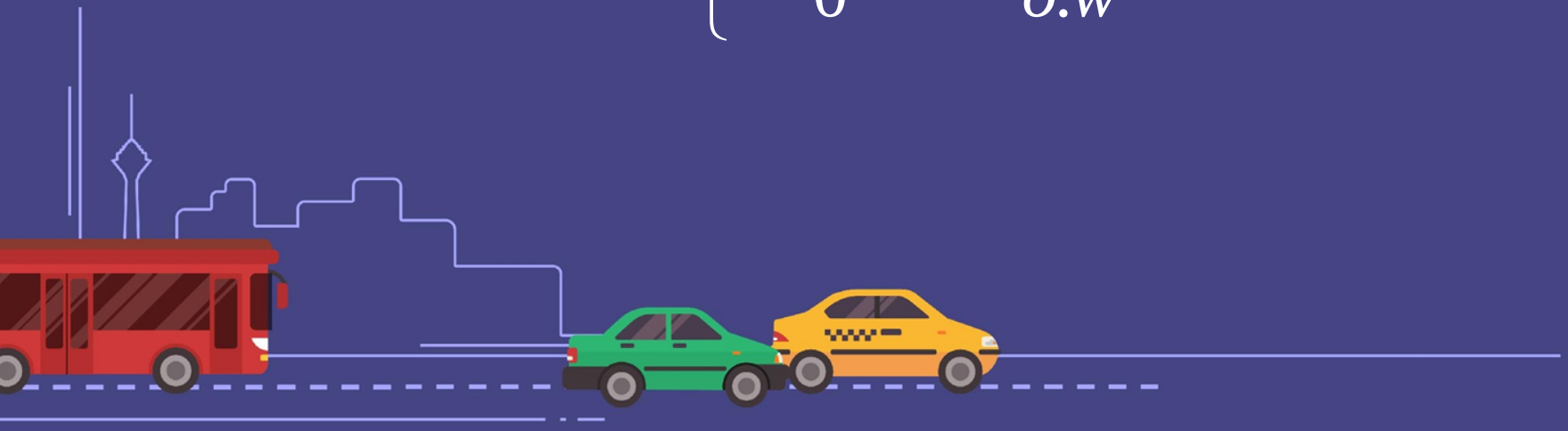
$$\mathrm{X}_t(r), \mathrm{X}_t(0) = \mathrm{X}$$



BOCD

Hazard:
let's assume that the 'natural chance' of a change point is $1/\lambda$

$$P(r_{t+1} = m \mid r_t = n) = \begin{cases} 1/\lambda & ,m = 0 \\ 1 - 1/\lambda & ,m = n+1 \\ 0 & o.w \end{cases}$$



BOCD

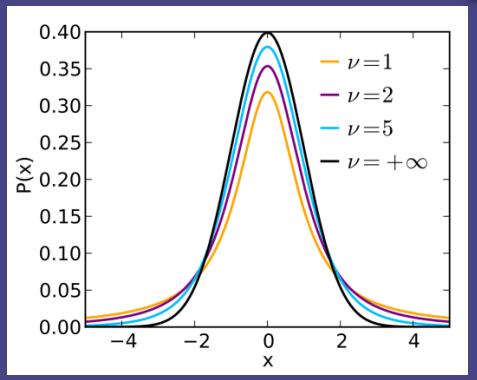
$$
\begin{aligned}
P(r_t, x_{1:t}) &= \sum_{r_{t-1}} P(r_t, r_{t-1}, x_{1:t}) \\
&= \sum_{r_{t-1}} P(r_t, x_t \mid r_{t-1}, x_{1:t-1}) P(r_{t-1}, x_{1:t-1}) \\
&= \sum_{r_{t-1}} P(r_t \mid r_{t-1}) P(x_t \mid r_{t-1}, x_{1:t}) P(r_{t-1}, x_{1:t-1})
\end{aligned}
$$

BOCD

$$
\begin{aligned}
P(r_t, x_{1:t}) &= \sum_{r_{t-1}} P(r_t, r_{t-1}, x_{1:t}) \\
&= \sum_{r_{t-1}} P(r_t, x_t \mid r_{t-1}, x_{1:t-1}) P(r_{t-1}, x_{1:t-1}) \\
&= \sum_{r_{t-1}} P(r_t \mid r_{t-1}) P(x_t \mid r_{t-1}, x_{1:t}) P(r_{t-1}, x_{1:t-1})
\end{aligned}
$$

BOCD

$$
\begin{aligned}
P(r_t, x_{1:t}) &= \sum_{r_{t-1}} P(r_t, r_{t-1}, x_{1:t}) \\
&= \sum_{r_{t-1}} P(r_t, x_t \mid r_{t-1}, x_{1:t-1}) P(r_{t-1}, x_{1:t-1}) \\
&= \sum_{r_{t-1}} P(r_t \mid r_{t-1}) P(x_t \mid r_{t-1}, x_{1:t}) P(r_{t-1}, x_{1:t-1})
\end{aligned}
$$



μ_0, ν, α, β	$\frac{\nu\mu_0 + n\bar{x}}{\nu+n}$, $\nu+n$, $\alpha+\frac{n}{2}$, $\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}$ • $\bar{x}$ is the sample mean

BOCD

$$P(r_t, x_{1:t}) = \sum_{r_{t-1}} P(r_t, r_{t-1}, x_{1:t})$$

$$= \sum_{r_{t-1}} P(r_t, x_t \mid r_{t-1}, x_{1:t-1}) P(r_{t-1}, x_{1:t-1})$$

$$= \sum_{r_{t-1}} P(r_t \mid r_{t-1}) P(x_t \mid r_{t-1}, x_{1:t}) P(r_{t-1}, x_{1:t-1})$$

BOCD

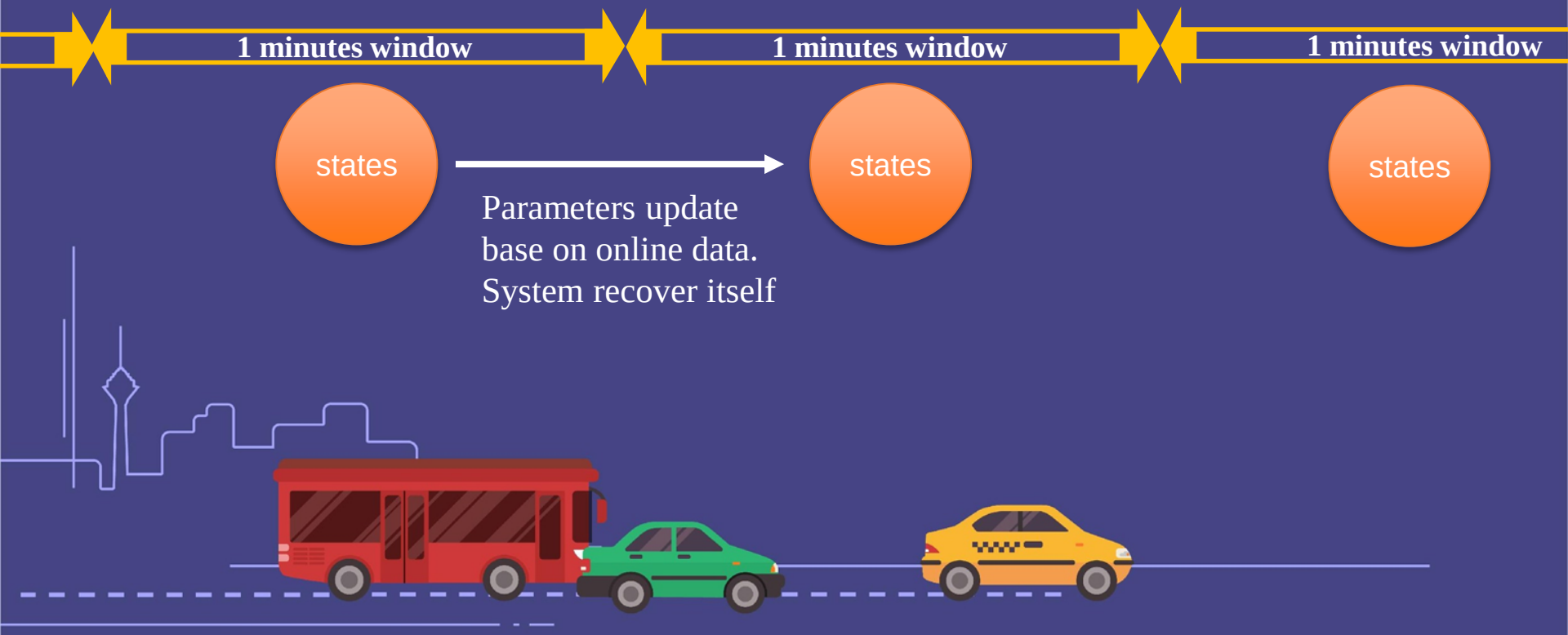
Bayesian online change point detection

$$r_t(n) = P(r_t \mid x_{1:t}) = \frac{P(r_t, x_{1:t})}{p(x_{1:t})}$$

Stateful Stream Processing

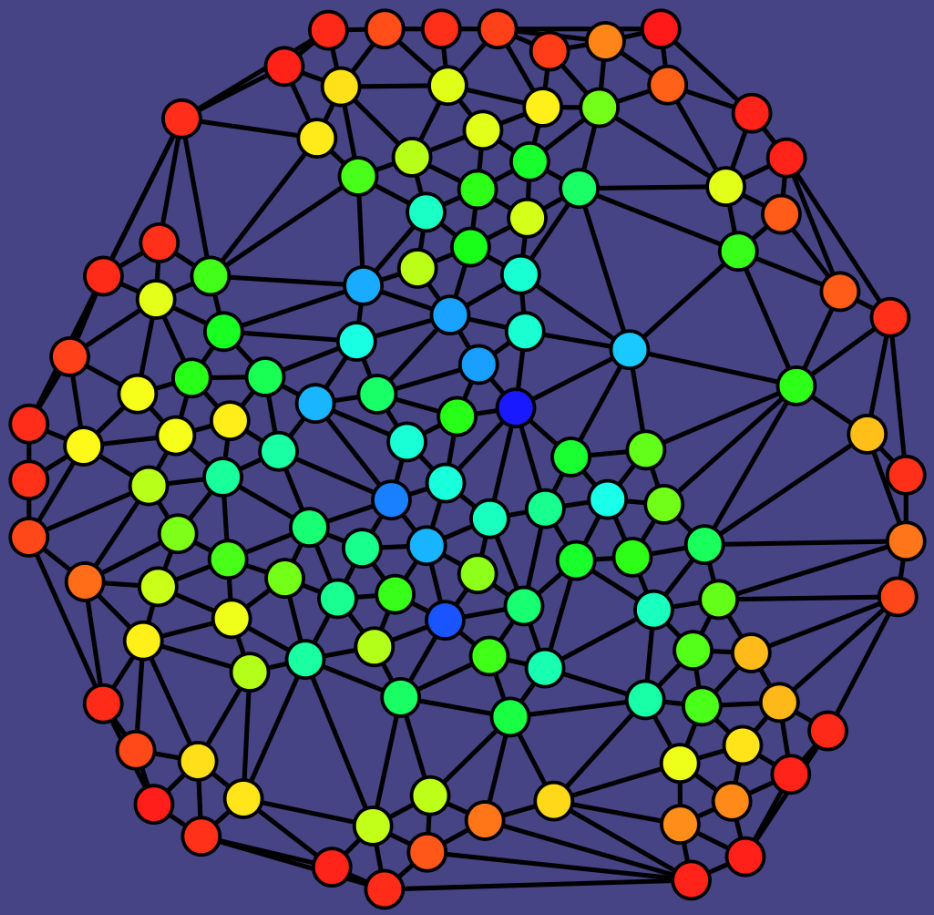
Online algorithms based on state

- Can use Flink for production
- Last 10 minutes of events are state
- Parameters of change point detection are state



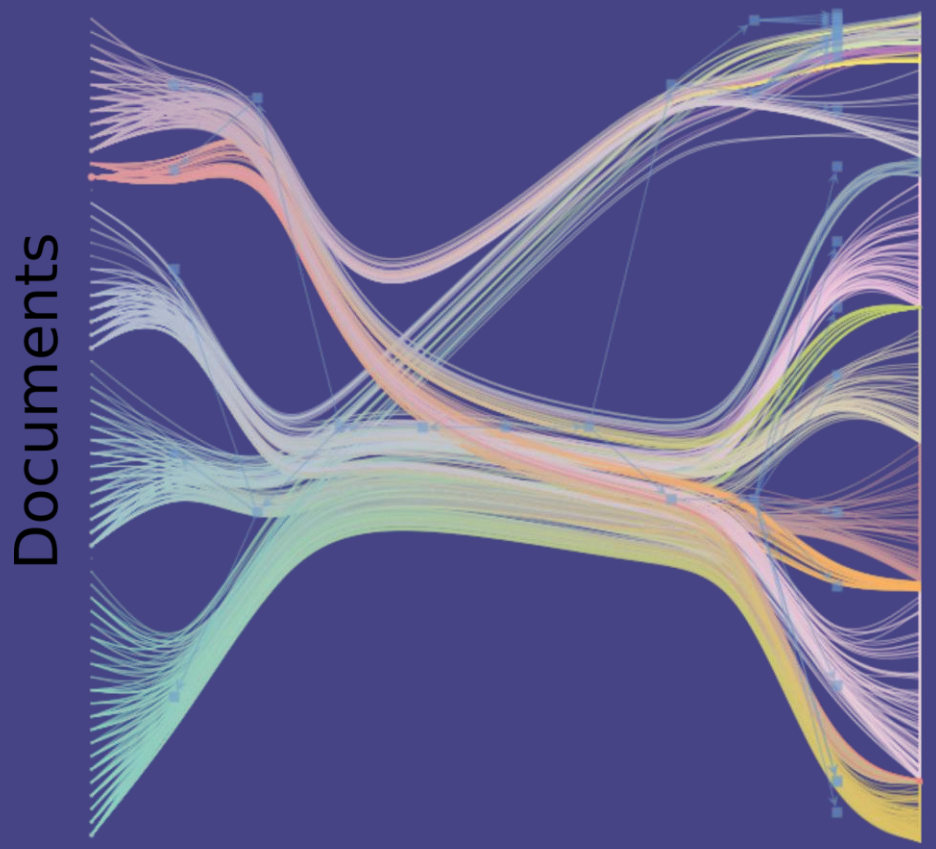
Future Work

- **Graph neural networks (GNNs)**
- **Road ranking**



Future Work

NLP-Topic modeling
Each road can be a document and
Events can be it’s words.



Questions