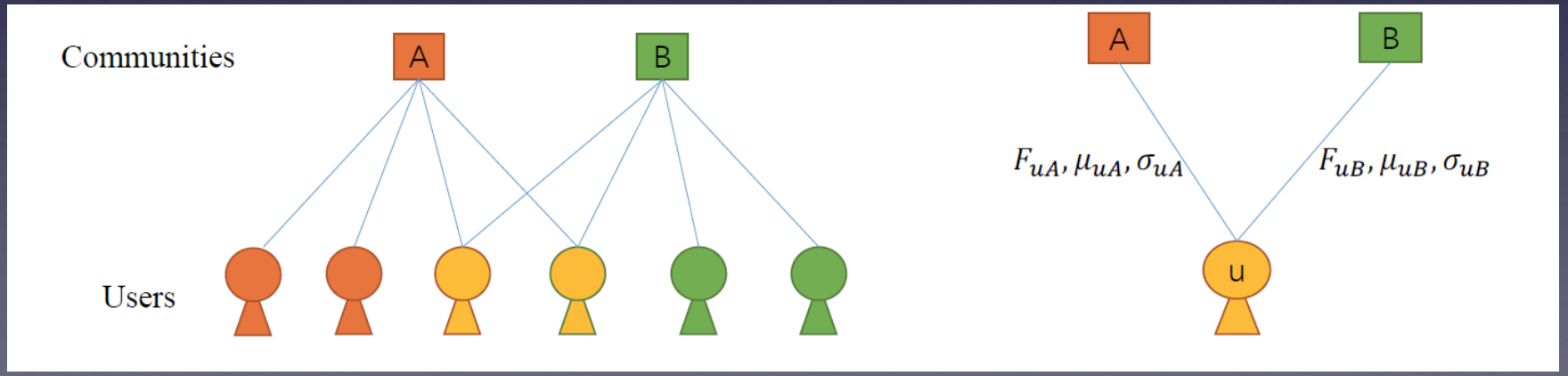
Dynamic Community Detection with Normal Distribution in Temporal Social Network

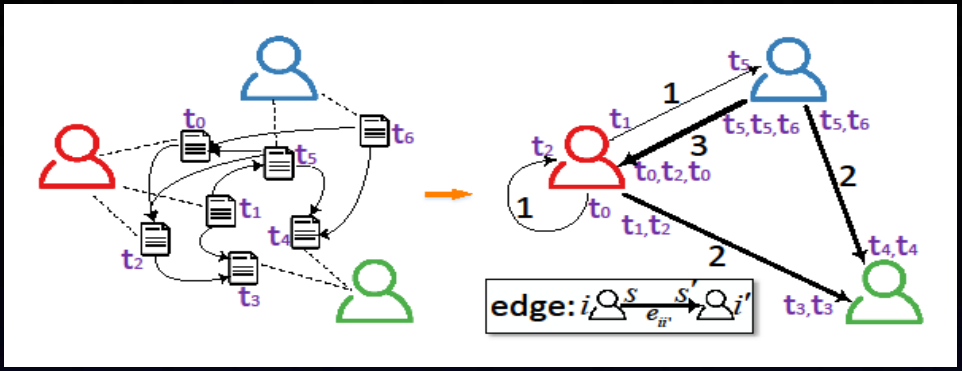
-Daniel (Huang Yaowei)

May 26

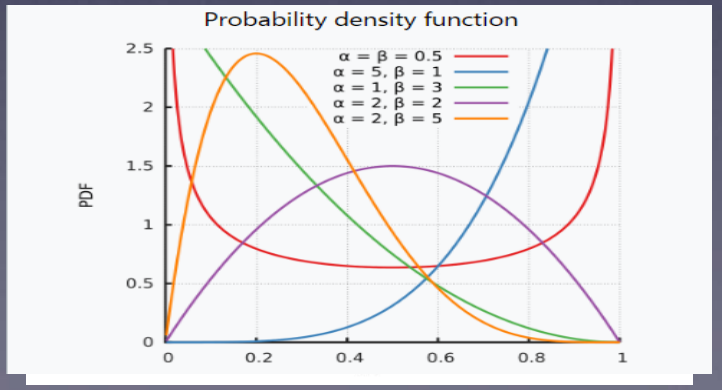
CDOT

- Each user will connect communities with a weight like F_{uA} and a Normal distribution with mean value μ_{uA} and variance σ_{uA}

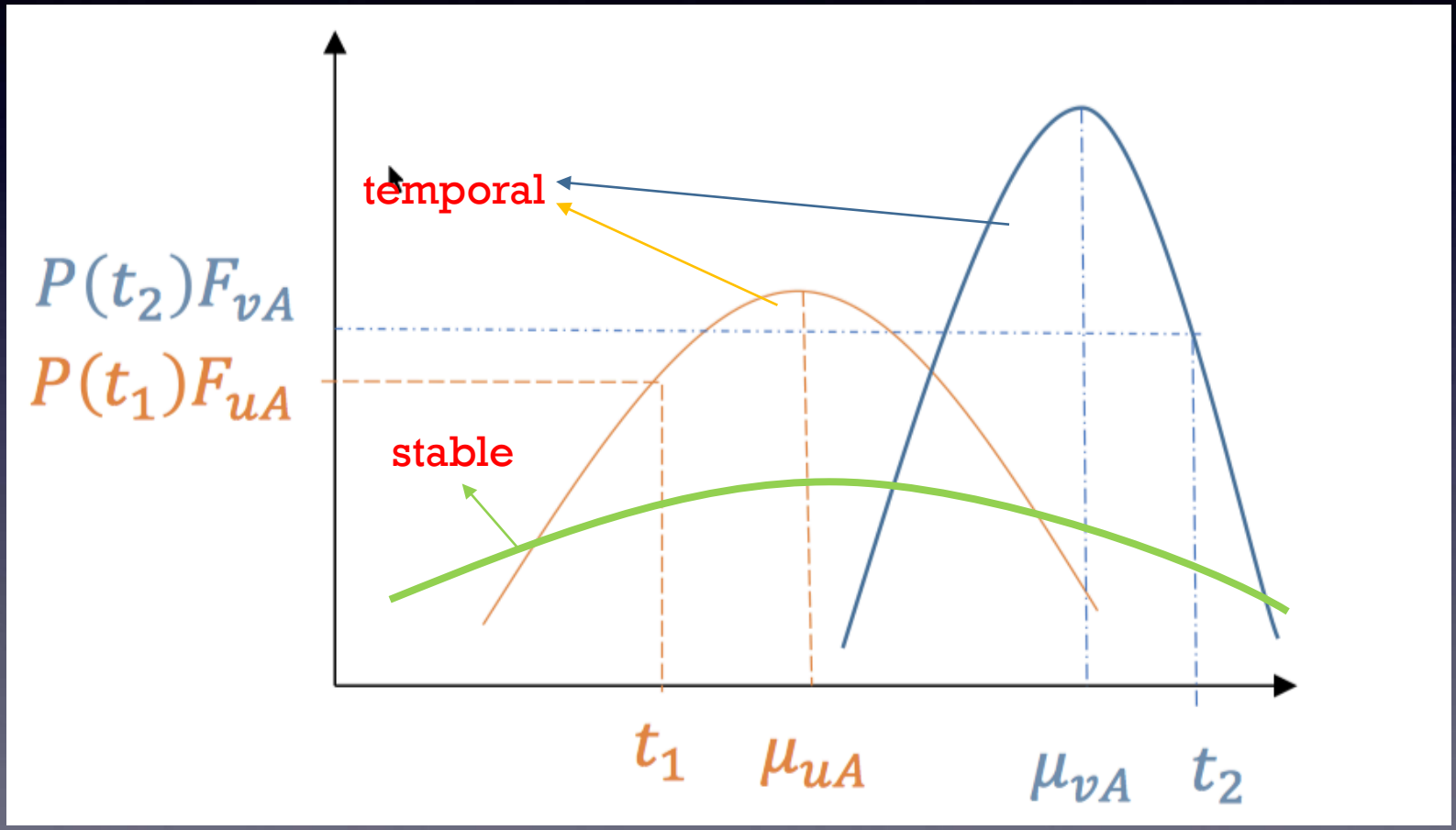




- $p(u,t_1,v,t_2) = 1-\exp(-\sum_c F_{uc} \cdot P_{uc}(t_1) \cdot F_{vc} \cdot P_{vc}(t_2))$
- $P_{uc}(t_1) = \frac{1}{\sqrt{2\pi}\sigma_{uc}} exp(-\frac{(t_1-\mu_{uc})^2}{2\sigma_{uc}^2})$
- **Why use Normal Distribution?**
  - **Compared with time slice(windows) -> less parameter**
  - **The shape is fixed -> convex optimization problem**
  - **The parameters are meaningful**



The parameters are meaningful



Parameter Learning

- $l(F, \mu, \sigma) = log\ P(G|F, \mu, \sigma)$:
  - $\hat{F}, \hat{\mu}, \hat{\sigma} = \underset{F \geq 0, \sigma > 0}{arg\max}\, l(F, \mu, \sigma)$
- Where
  - $l(F, \mu, \sigma) = \sum_{(u,t_1,v,t_2) \in E} log(1 - exp(-H) - \sum_{(u,t_1 v,t_2) not \in E} H$
- where
  - $H = \sum_C F_{\text{uc}} \cdot \mathbf{P}_{uc}(t_1) \cdot F_{vc} \cdot \mathbf{P}_{vc}(t_2)$

Parameter Learning

- $l(F_u) = \sum_{(u,t_1,v,t_2)\in N(u)} log(1 - exp(-H) - \sum_{(u,t_1,v,t_2)\in Neg\ N(u)} H$
- Projected gradient ascent
  - $F_{uc}^{new} = \max(0, F_{uc}^{old} + \alpha_{F_u} \nabla l(F_{uc}))$
- Where
  - $\nabla l(F_{uc})) = \sum_{(u,t_1,v,t_2)\in N(u)} \frac{exp(-H)}{1-exp(-H)} P_{vc}(t_2) F_{vc} P_{uc}(t_1) - \sum_{(u,t_1 v,t_2)\in Neg\ N(u)} P_{vc}(t_2) F_{vc} P_{uc}(t_1)$

$$\nabla l(\mu_{uc}) = \sum_{(v,t_2)\in N(u)} \frac{\exp(-H)}{1-\exp(-H)} F_{uc} P_{vc}(t_2) F_{vc} P_{uc}(t_1) \frac{t_1 - \mu_{uc}}{\sigma_{uc}^2} - \sum_{(v,t_2)\in Neg N(u)} F_{uc} P_{uc}(t_1) F_{vc} P_{vc}(t_2) \frac{t_1 - \mu_{uc}}{\sigma_{uc}^2} \quad (9)$$

$$\nabla l(\sigma_{uc}) = \sum_{(v,t_2)\in N(u)} \frac{\exp(-H)}{1-\exp(-H)} F_{uc} P_{vc}(t_2) F_{vc} P_{uc}(t_1) \frac{(t_1 - t_{uc})^2 - \sigma_{uc}^2}{\sigma_{uc}^3} - \sum_{(v,t_2)\in Neg N(u)} F_{uc} P_{uc}(t_1) F_{vc} P_{vc}(t_2) \frac{(t_1 - t_{uc})^2 - \sigma_{uc}^2}{\sigma_{uc}^3} \quad (11)$$

Negative Edges Sampling

- First reason: the Negative edges is infinity because time t is a continuous distribution.

- Second reason: in a huge network a node will only have some edges between several nodes which are close to it or having high weight in the same community, not all node.

$$F_{uc} \longrightarrow 0$$

Determing Community Membership

- Add a threshold

$$\delta_k = \sqrt{-\log(1 - \frac{2|E|}{|V|(|V|-1)})}$$

- "stable" user

$$\mathrm{F}_{\mathrm{uc}} P(\mu) > \delta_k$$ "daniu"

- "temporal" user

Usually belongs to the community "daniu"