Dynamic Community Detection with Normal Distribution in Temporal Social Networks

Yaowei Huang
5140309539
14330222150355@sjtu.edu.cn

Yuchen Lin
5140309507
yuchenlin@sjtu.edu.cn

Zhaozhe Song
5140309514
zhaozhesong@sjtu.edu.cn

Supervisor: Prof. Luoyi Fu

May 2017

Are you still close to your friends in high school?

Are you still close to your friends in high school?

No, we are closer to
our friends in university now.

Are you still close to your friends in high school?

No, we are closer to
our friends in university now.

$\boxed{\text{Connections amongst people are dynamic.}}$

Are you still close to your friends in high school?

No, we are closer to
our friends in university now.

Connections amongst people are dynamic.

We wanna measure the dynamic connections between people and communities.

Outline

Outline

- Background of Community Detection

Outline

- Background of Community Detection
- Motivation of Dynamic Community Detection

Outline

- Background of Community Detection
- Motivation of Dynamic Community Detection
- Observation and Intuition

Outline

- Background of Community Detection
- Motivation of Dynamic Community Detection
- Observation and Intuition
- Brief Introduction of Our Approach

Outline

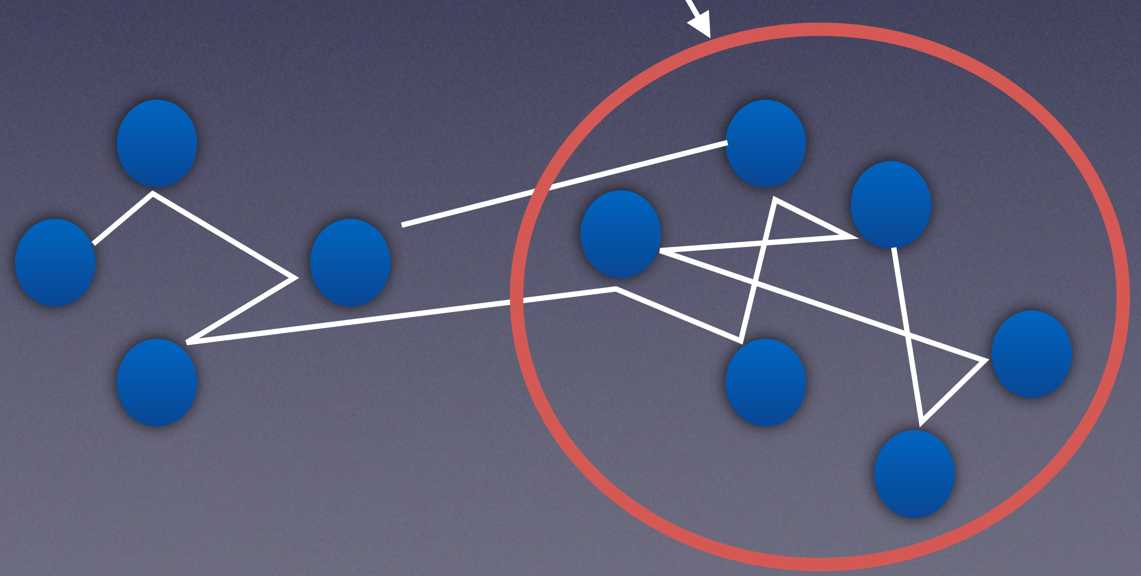
- Background of Community Detection
- Motivation of Dynamic Community Detection
- Observation and Intuition
- Brief Introduction of Our Approach
- Challenges

Background

Definition: A **community** is a group of nodes where the connections amongst them are much denser than their connections with outside nodes.

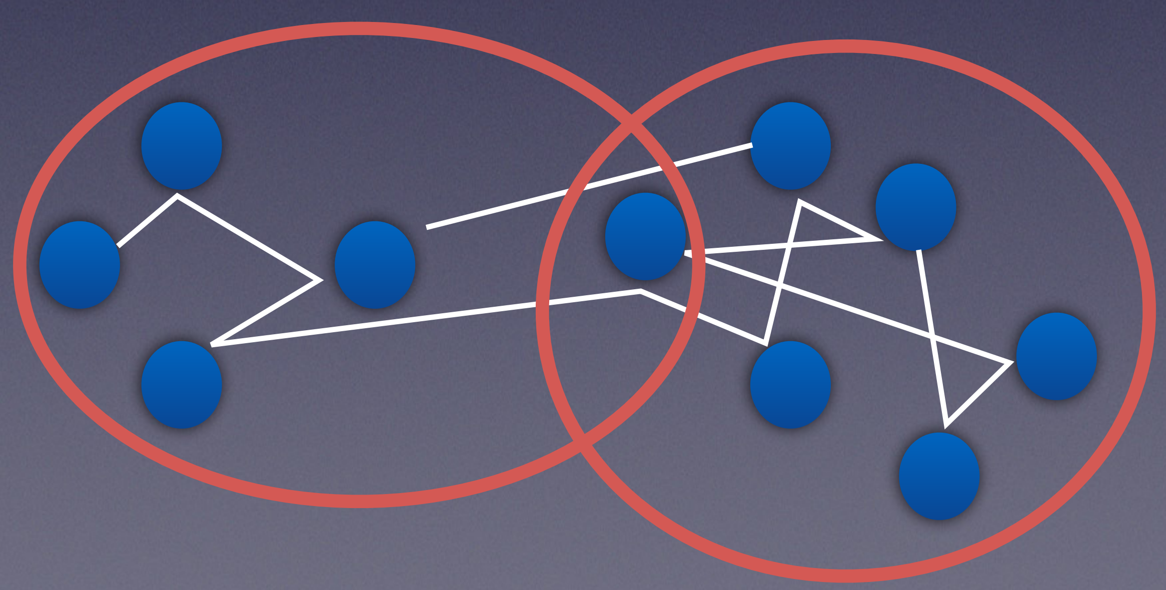
Background

Definition: A **community** is a group of nodes where the connections amongst them are much denser than their connections with outside nodes.



Background

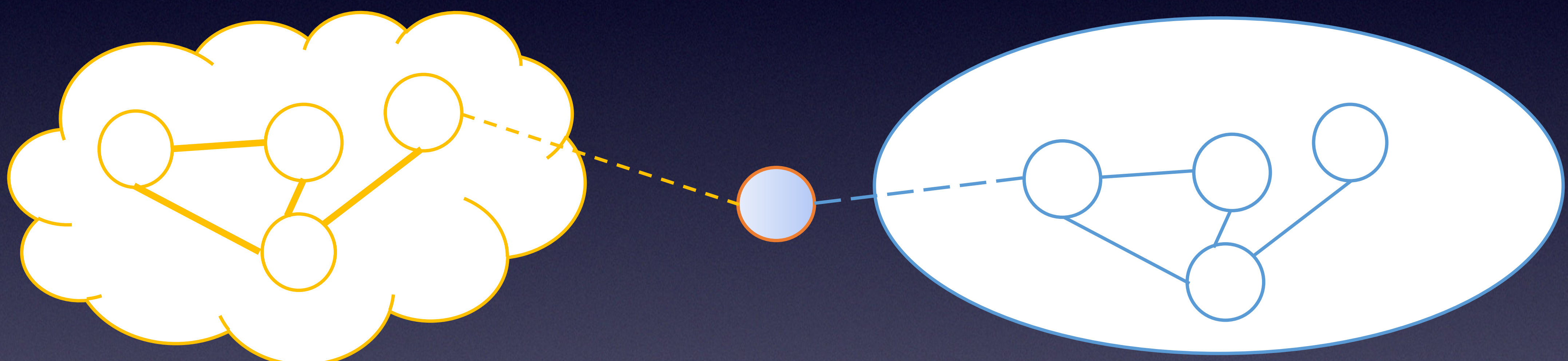
Definition: An **overlapping community** allows some nodes belong to multiple communities.



Static community

Static community

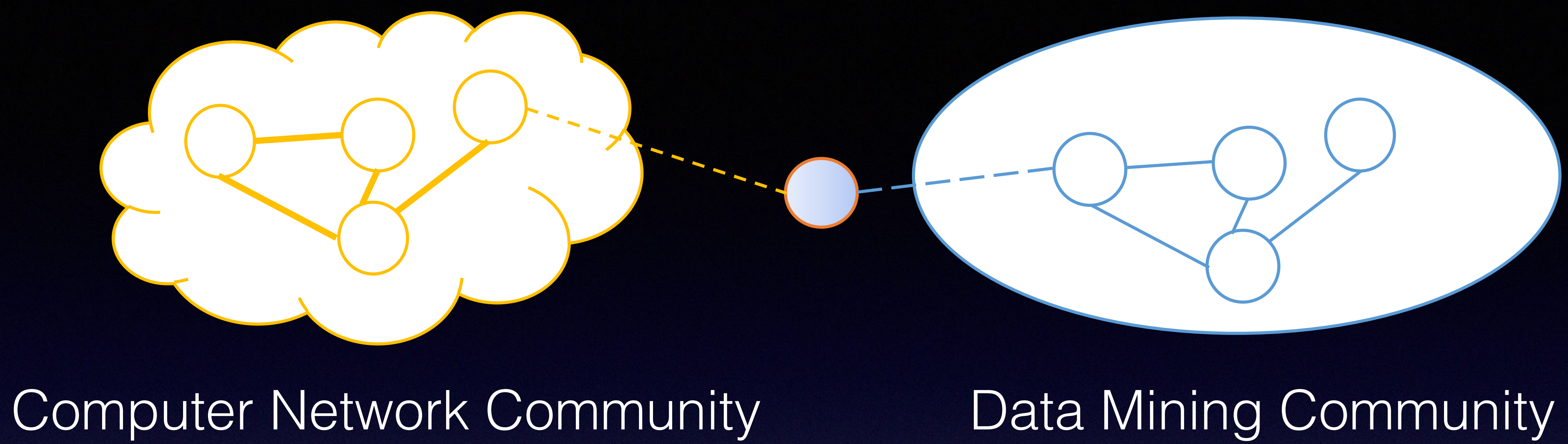
An edge **(u,v)** denotes an interaction between nodes



Computer Network Community

Data Mining Community

Example: A static collaboration network within a long period of time (from being Master student to PhD student)



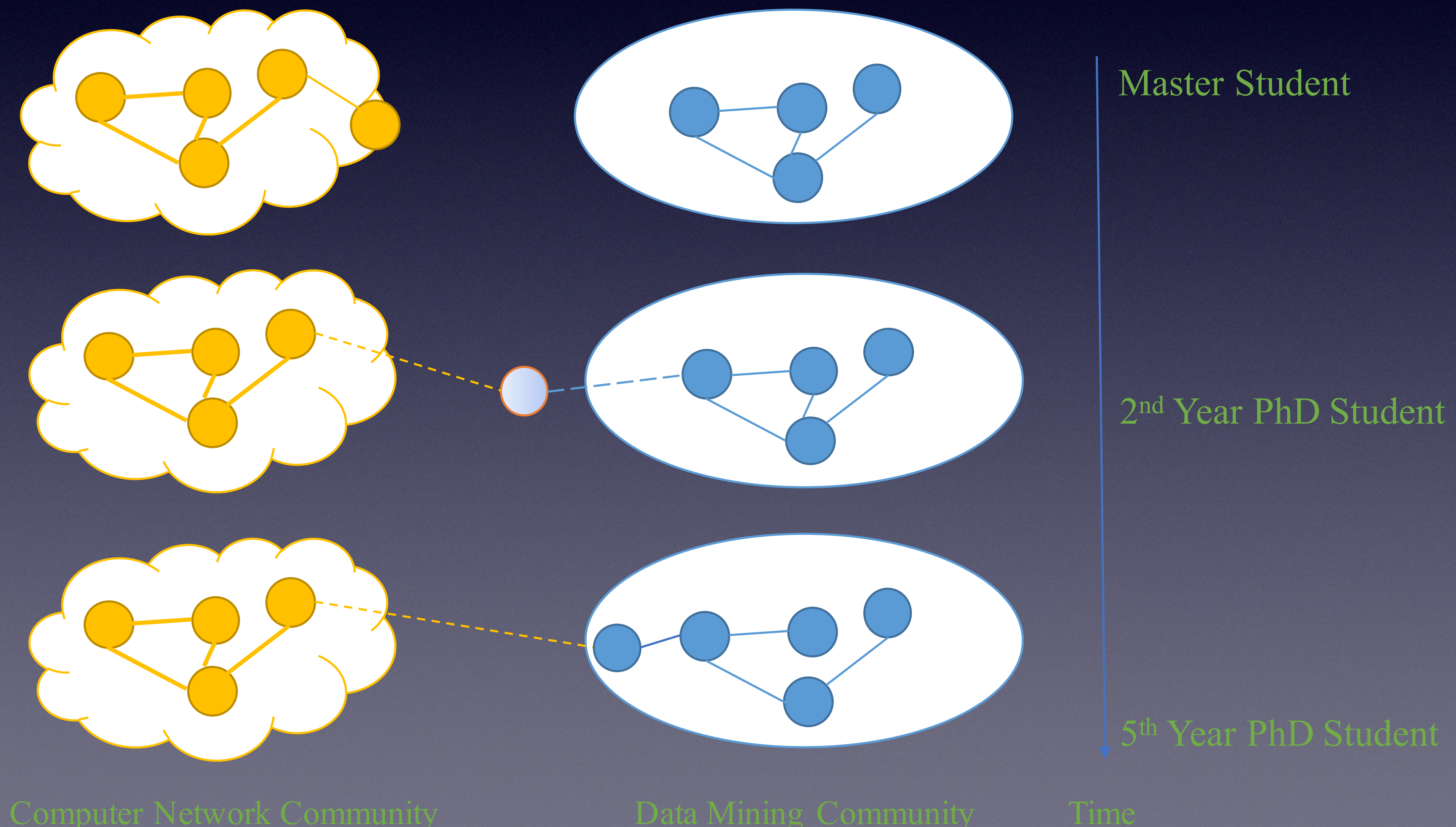
Three Possibilities from This Figure:

- this guy shifted from Computer Network to Data Mining
- this guy shifted from Data Mining to Computer Network
- this guy always did research both of them

What’s the truth?

Dynamic community in dynamic (temporal) networks

An edge **(u, v, t_u, t_v)** denotes
an interaction and its related times between nodes



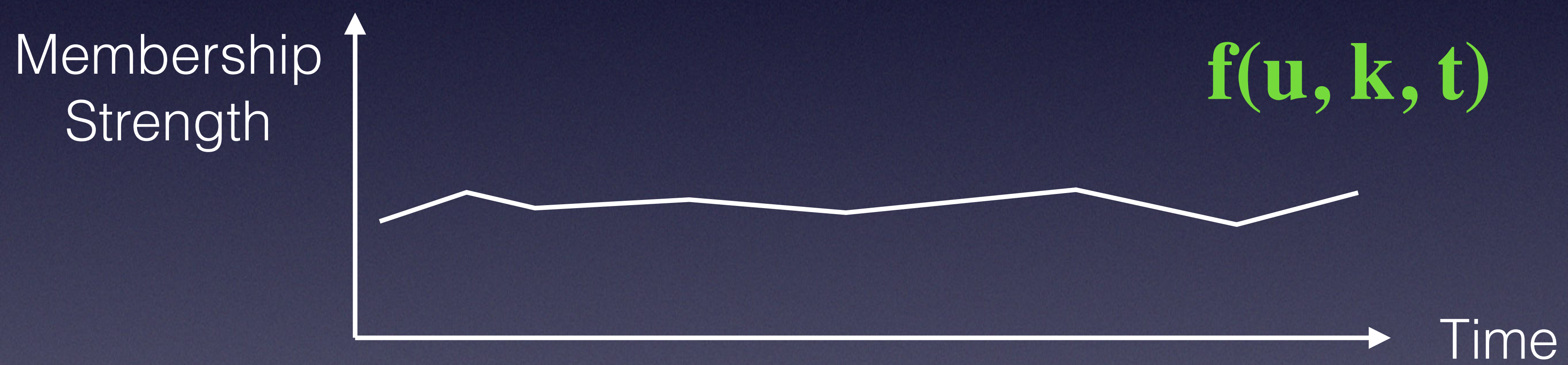
Observation

Most membership strength in a temporal social networks are:

Observation

Most membership strength in a temporal social networks are:

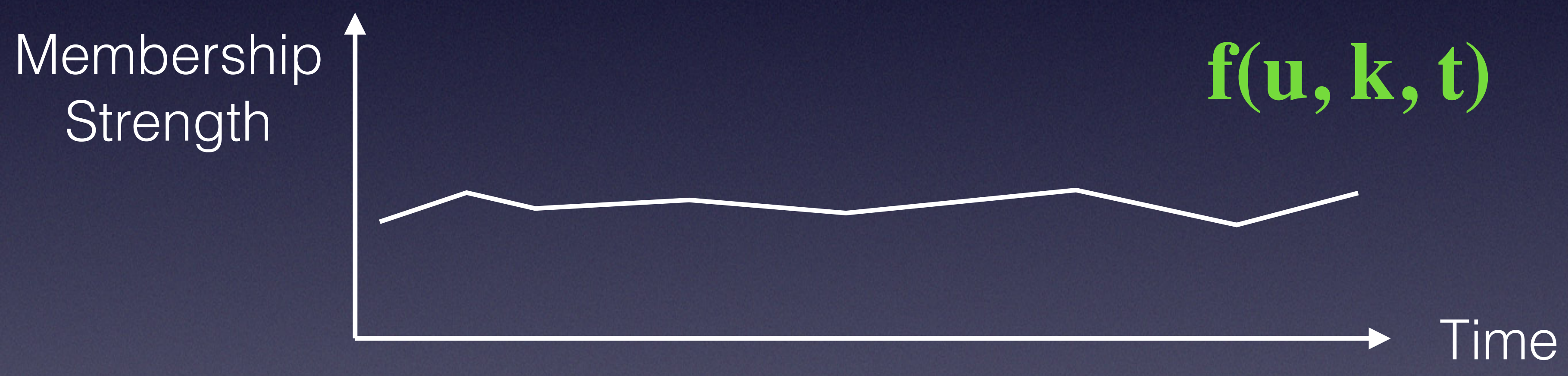
1) stable, not time-varying



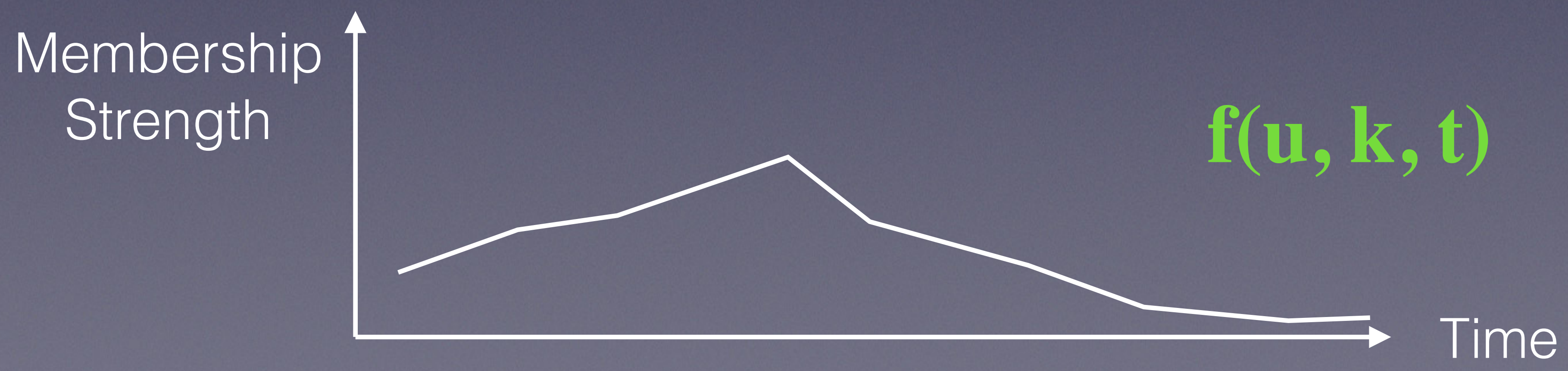
Observation

Most membership strength in a temporal social networks are:

1) stable, not time-varying

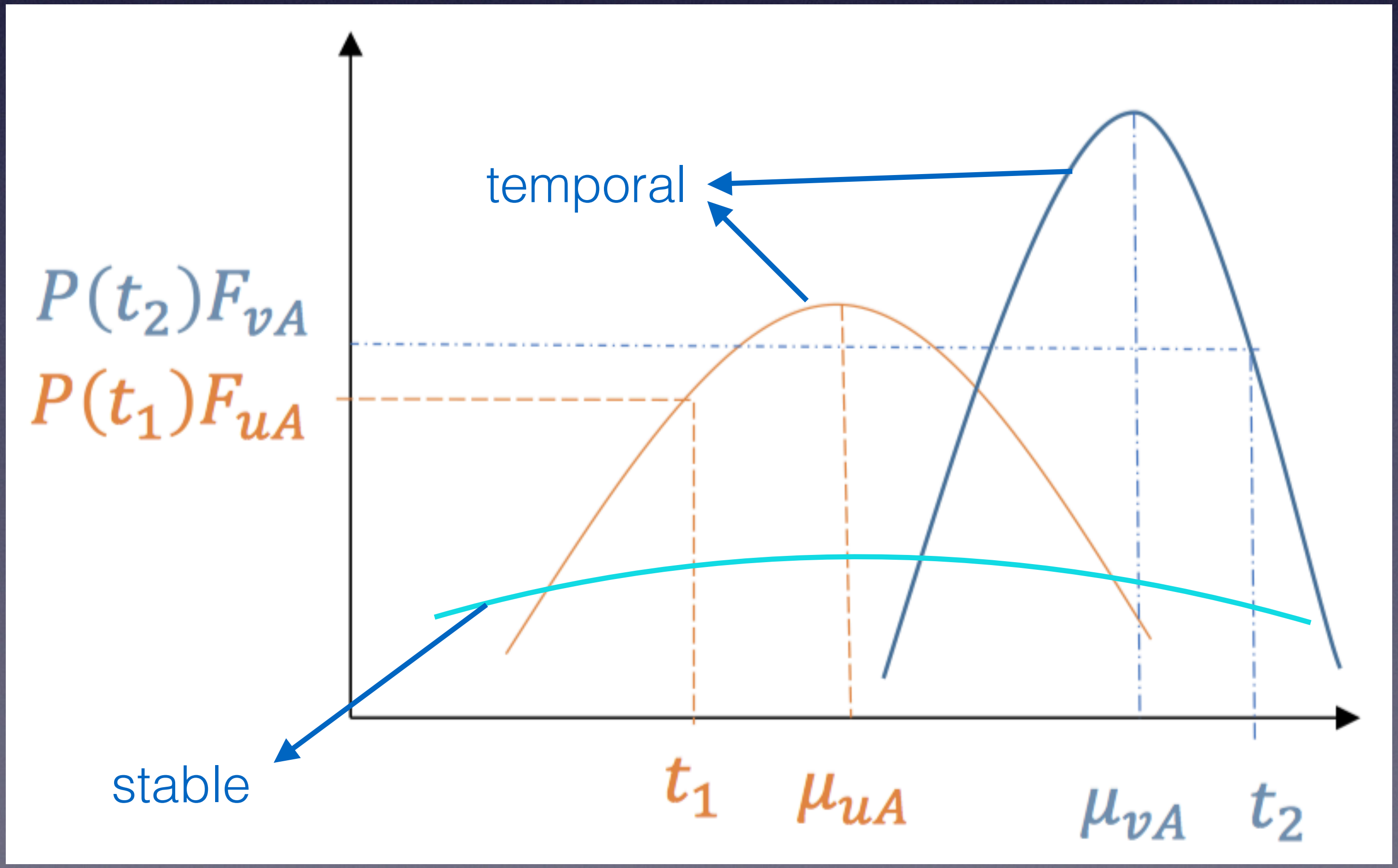


2) temporal and always with mono peak



Our Intuition

Use Normal Distribution to model the strength of membership — adding a pair of (μ , σ) as temporal factors to describe strength of membership



Problem Definition: Community Detection

Problem Definition: Community Detection

- **Given n nodes (people) and m edges (interaction) -> G(V,E)**
- **To find number of k communities where people belong to**
- **TWO observation:**
  - If two people interact often -> P (same community)
  - If two people belong to same community -> P (Interact)

Modeling for static community detection in BigClam

- Define:
  - F_{uc} -> weight between user u and community c
  - P(u, v) -> The possibility of interaction between users u and v
- Assume one community c connects users u and v with

$$P(u,v|c) = 1-\exp(-F_{uc} \cdot F_{vc})$$

- We have

$$P(u,v) = 1-\exp(-\sum_{C} F_{\mathrm{uc}} \cdot F_{vc})$$

Learning for static community detection in BigClam

How can we get F:

- **Maximizing the log likelihood of:**

$$\boldsymbol{l(F) = log\ P(G|F)}$$

$$\hat{\mathrm{F}} = \underset{\boldsymbol{F \geq 0}}{\boldsymbol{arg}\max}\, \boldsymbol{l(F)}$$

- **where**

$$l(F) = \sum_{(u,v)\in E} \log(1 - \exp(-F_u F_v^T)) - \sum_{(u,v)\notin E} F_u F_v^T$$

Learning for static community detection in BigClam

Block coordinate gradient ascent algorithm

$$l(F) = \sum_{(u,v)\in E} \log(1-\exp(-F_u F_v^T)) - \sum_{(u,v)\notin E} F_u F_v^T$$

We can solve subproblem for each u,

- **Update F_u for each u with the other F_v fixed**
- **Became a convex optimization problem**

$$l(F_u) = \sum_{v\in\mathcal{N}(u)} \log(1-\exp(-F_u F_v^T)) - \sum_{v\notin\mathcal{N}(u)} F_u F_v^T$$

$$\nabla l(F_u) = \sum_{v\in\mathcal{N}(u)} F_v \frac{\exp(-F_u F_v^T)}{1-\exp(-F_u F_v^T)} - \sum_{v\notin\mathcal{N}(u)} F_v$$

8 /15

Our Modeling for Dynamic Community Detection

Assume every user has a time distribution between community c

- $\text{Gaussian}_{uc} \sim (\mu_{uc}, \sigma^2_{uc})$
- $P(u,v) = 1-\exp(-\sum_C F_{uc} \cdot F_{vc})$ ->

$$P(u,t_1,v,t_2) = 1-\exp(-\sum_C F_{uc} \cdot \text{Gau}_{uc}(t_1) \cdot F_{vc} \cdot \text{Gau}_{vc}(t_2))$$

Parameter Learning for Dynamic Community Detection

Maximizing the log likelihood of:

- $\boldsymbol{l(F, Gau) = log\ P(G|F, Gau)}$ **Gau -> f**

$$\hat{F}, \hat{f} = \underset{F \geq 0, f \geq 0}{arg\max}\, l(F, f) \qquad f(t_1|u,k) = \frac{1}{\sqrt{2\pi}\sigma_{uk}} \exp(-\frac{(t_1 - t_{uk})^2}{2\sigma_{uk}^2})$$

$$l(F, f) = \sum_{(u,v,t_1,t_2) \in E} log(1 - \exp(-\sum_{k=0}^{C} F_{uk} f(t_1|u,k) F_{vk} f(t_2|v,k)))$$
$$- \sum_{(u,v,t_1,t_2) \notin E} \sum_{k=0}^{C} F_{uk} f(t_1|u,k) F_{vk} f(t_2|v,k))$$

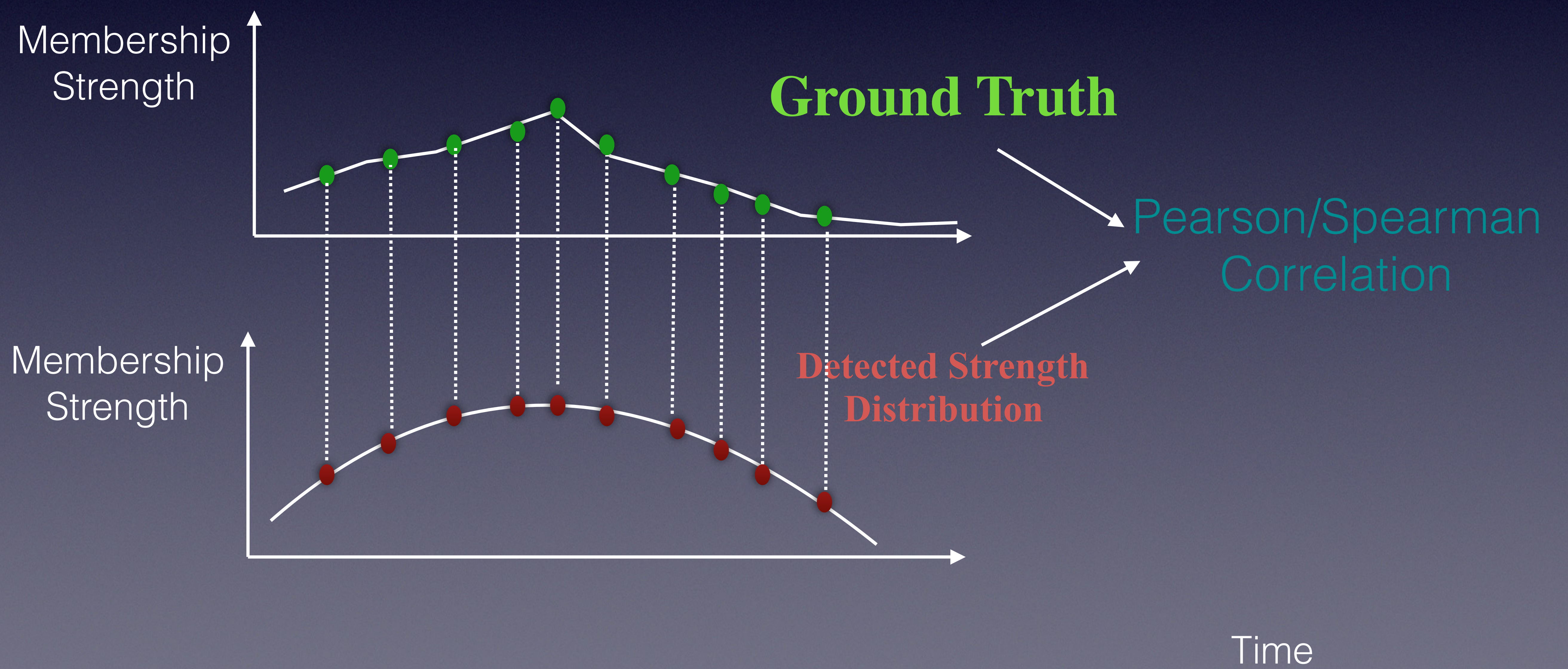
Random Sampling of Negative Samples

Evaluation on the membership strength matrix F

- Relative Accuracy
- Average F1 Score
- Omega Index
- Normalized Mutual Information

Evaluation on the estimated temporal factors (μ , σ)

- Pearson/Spearman Correlation



Thank You!