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

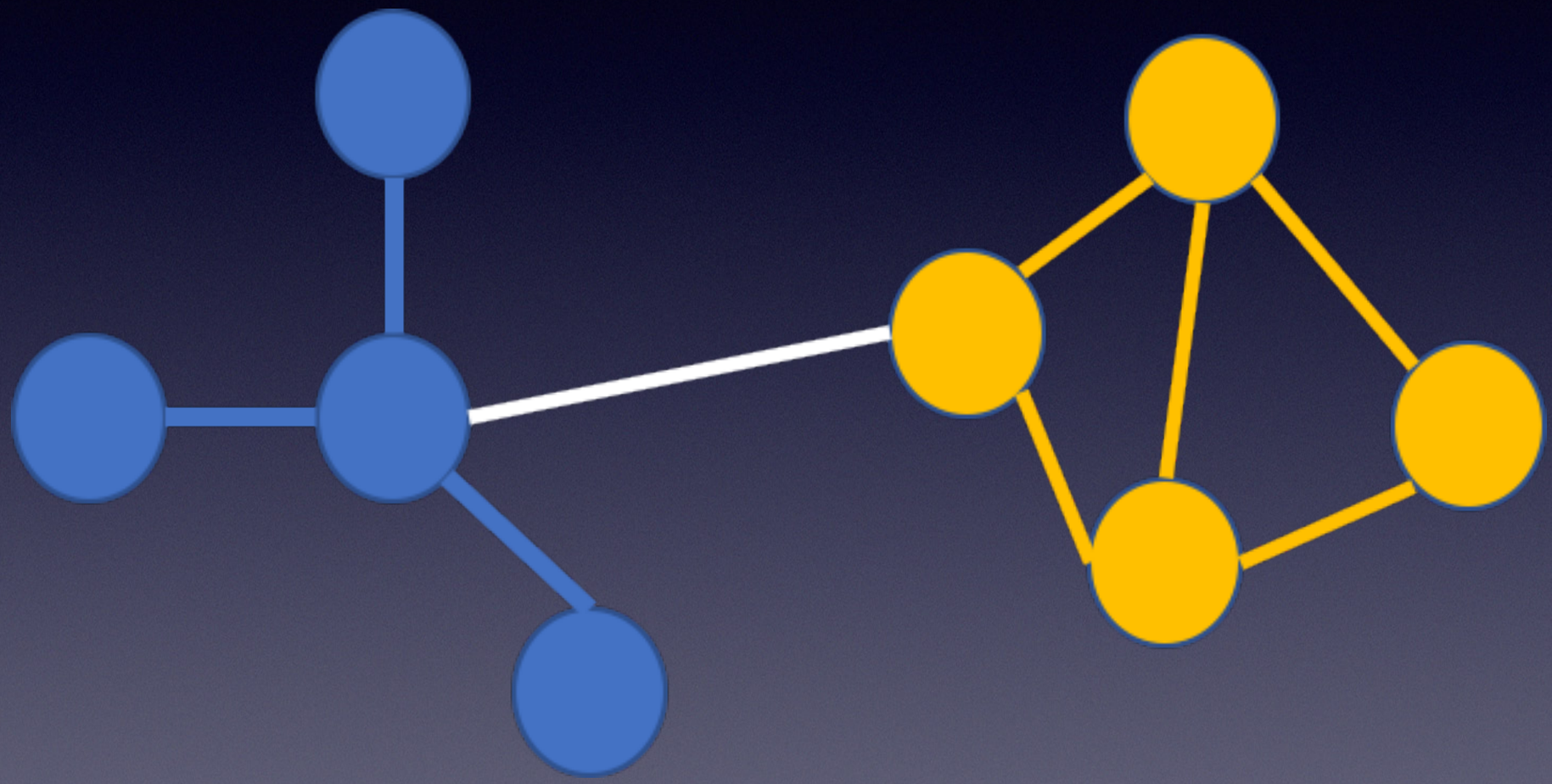
Evaluation and Simulation

宋肇哲

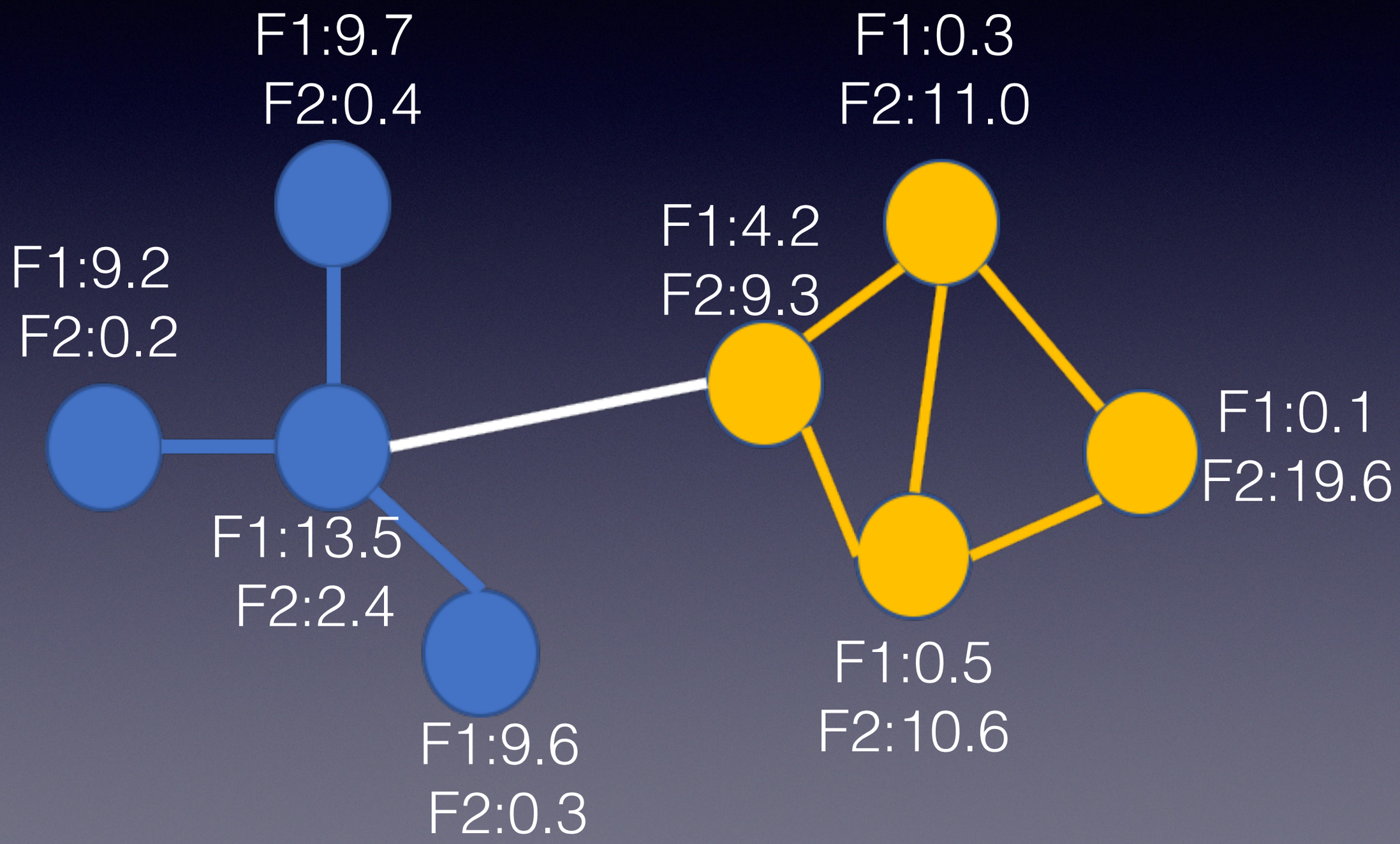
Evaluation and Simulation

Novel. Have to design some metrics by ourselves.

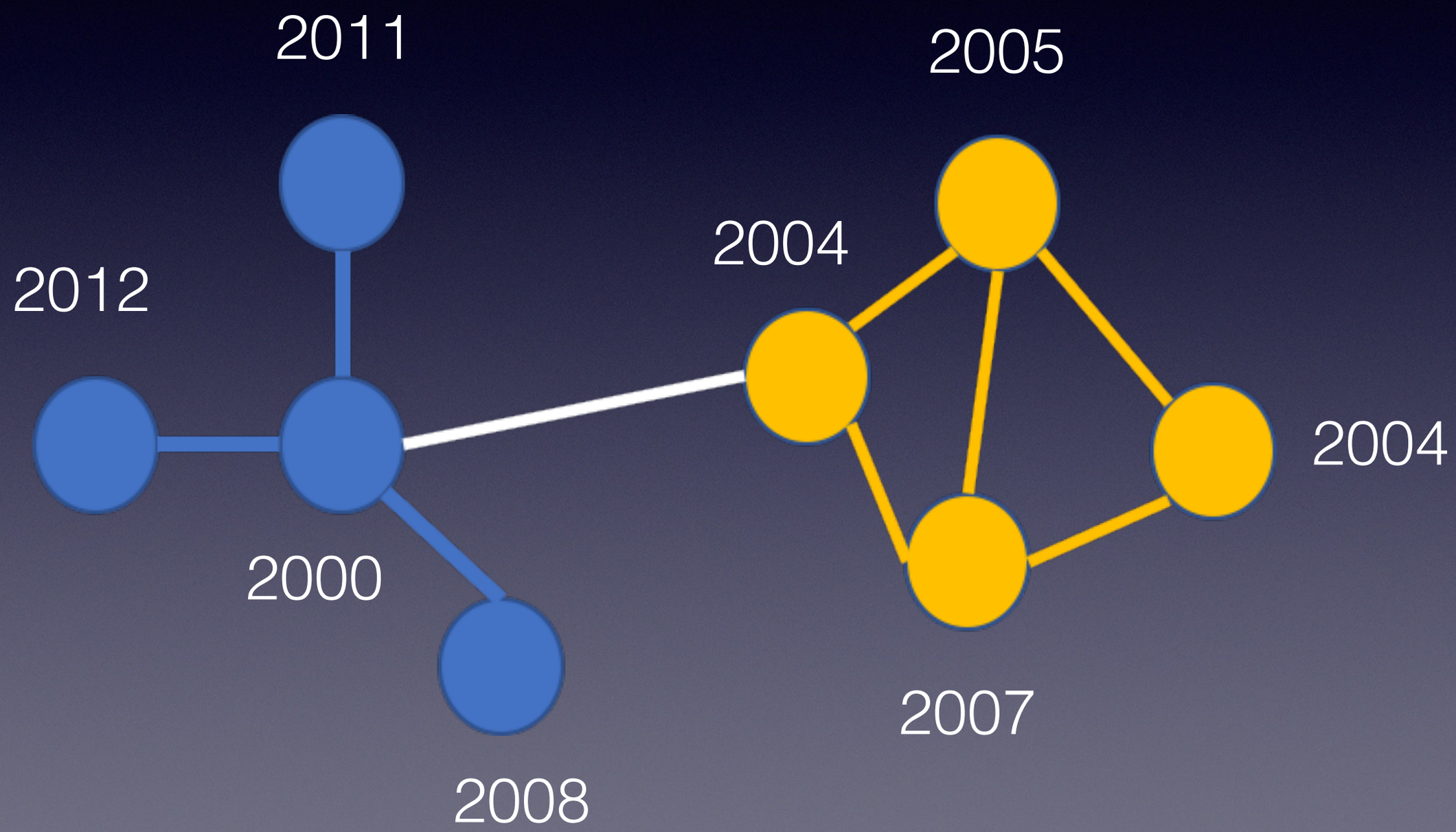
Example



Example Values of F



Example Values of μ



But, how can we evaluate the results **quantitatively**?

Two aspects

- The community weight (F)
- The temporal dimension (μ , σ)

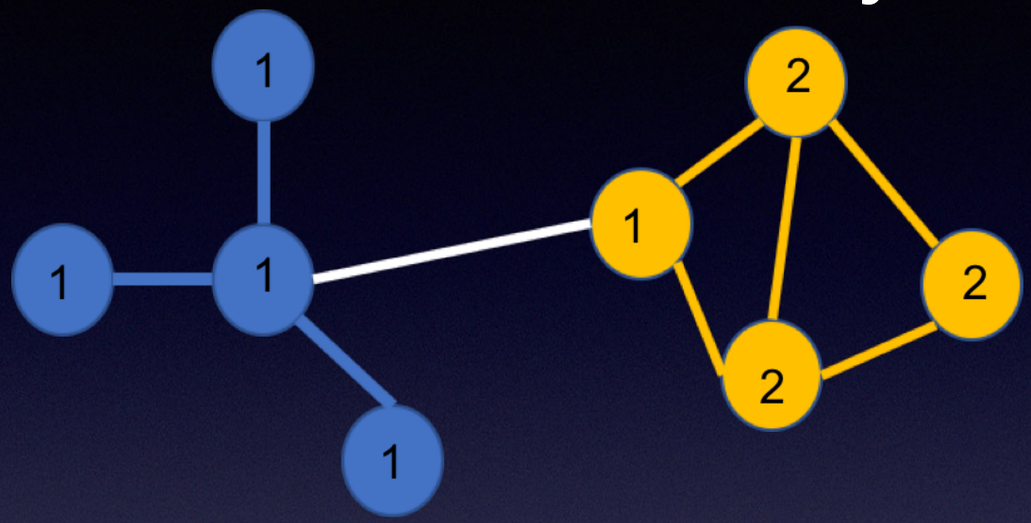
Evaluation on the community weight F

- Average F1 Score
- Omega index
- Accuracy in the number of communities

Problem:

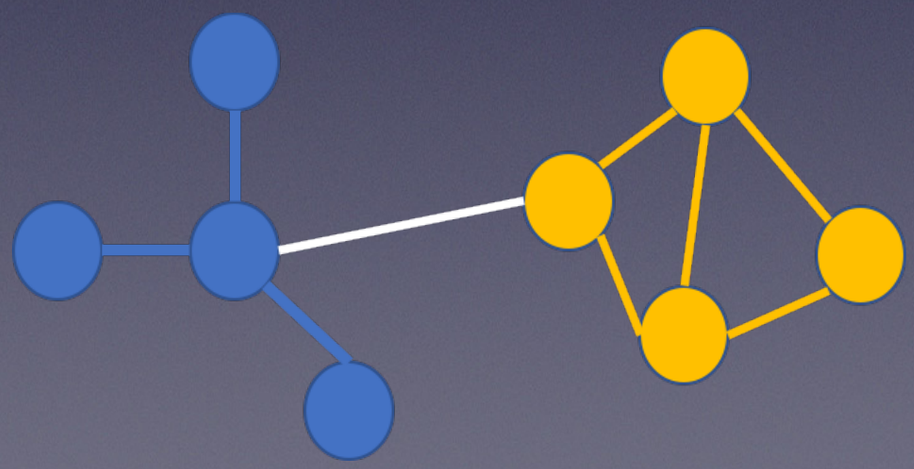
Ground truth: Community number

C^*



But our detected result….
Only anonymous communities

$\hat{C}$

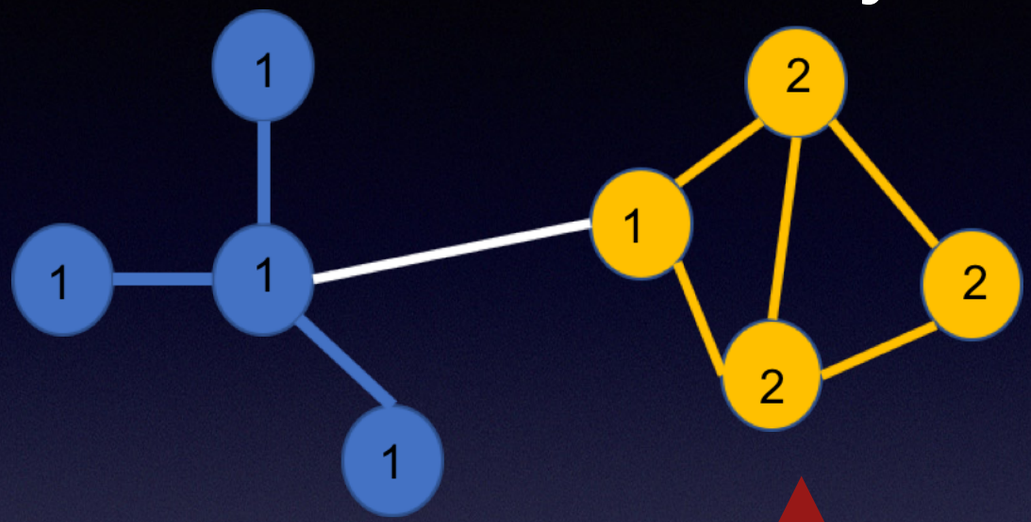


Find the most
similar matching for
each community!

Problem:

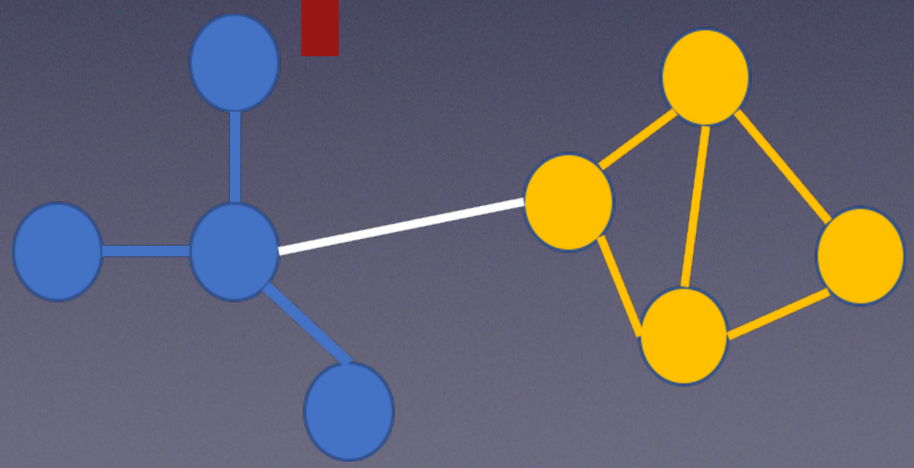
Ground truth: Community number

C^*

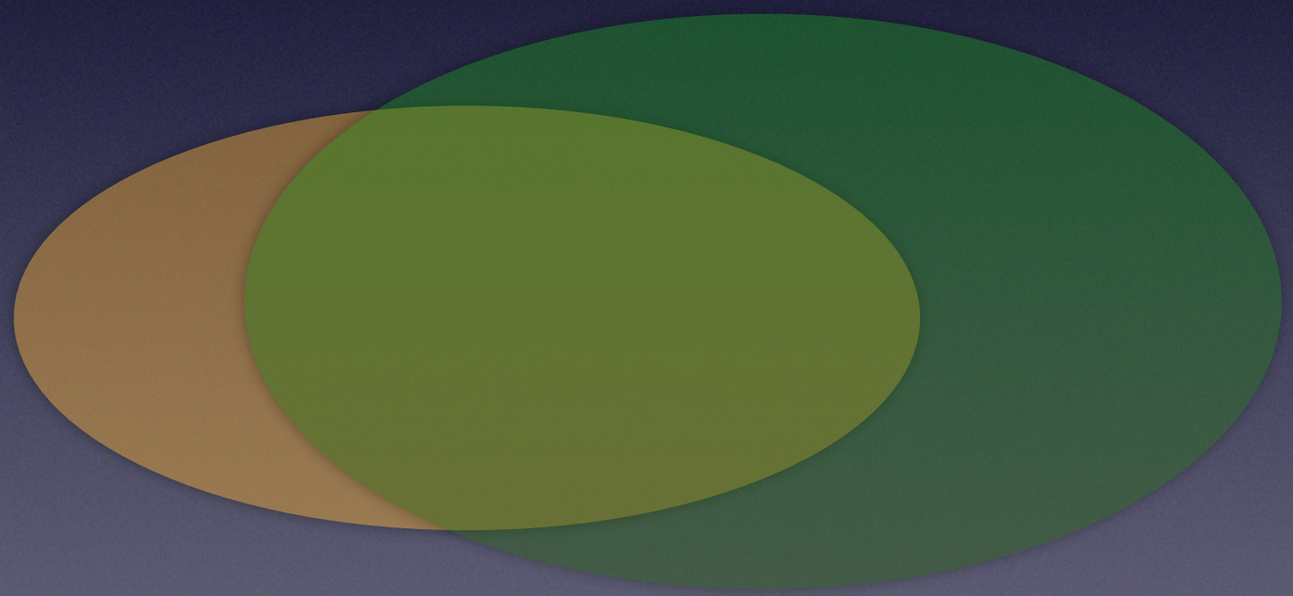


But our detected result….
Only anonymous communities

$\hat{C}$



$F1(C_i, \hat{C}_j)$ is the harmonic mean of Precision and Recall



Evaluation on the community weight F

- Average F1 Score
- Omega index
- Accuracy in the number of communities

Average over all detected and ground truth communities:

$$\frac{1}{2}\left(\frac{1}{|C^*|}\sum_{C_i \in C^*} F1(C_i, C_{\hat{g(i)}}) + \frac{1}{|\hat{C}|}\sum_{\hat{C}_i \in \hat{C}} F1(C_{g'(i)}, \hat{C}_i)\right)$$

where the best matching g and g' is defined as follows:

$$g(i) = \underset{j}{\operatorname{argmax}}\, F1(C_i, \hat{C}_j), \quad g'(i) = \underset{j}{\operatorname{argmax}}\, F1(C_j, \hat{C}_i)$$

The best matching for ground truth

The best matching for our detected result

*Note: not one-to-one matching

Evaluation on the community weight F

- Average F1 Score
- Omega index
- Accuracy in the number of communities

Omega index

$$\frac{1}{|V|^2} \sum_{u,v \in V} \mathbf{1}\{|C_{uv}| = |\hat{C_{uv}}|\}$$

estimating the number of communities that each pair of nodes shares

Evaluation on the community weight F

- Average F1 Score
- Omega index
- Accuracy in the number of communities

Accuracy in the number of communities

$$1 - \frac{||C^*| - |\hat{C}||}{2|C^*|}$$

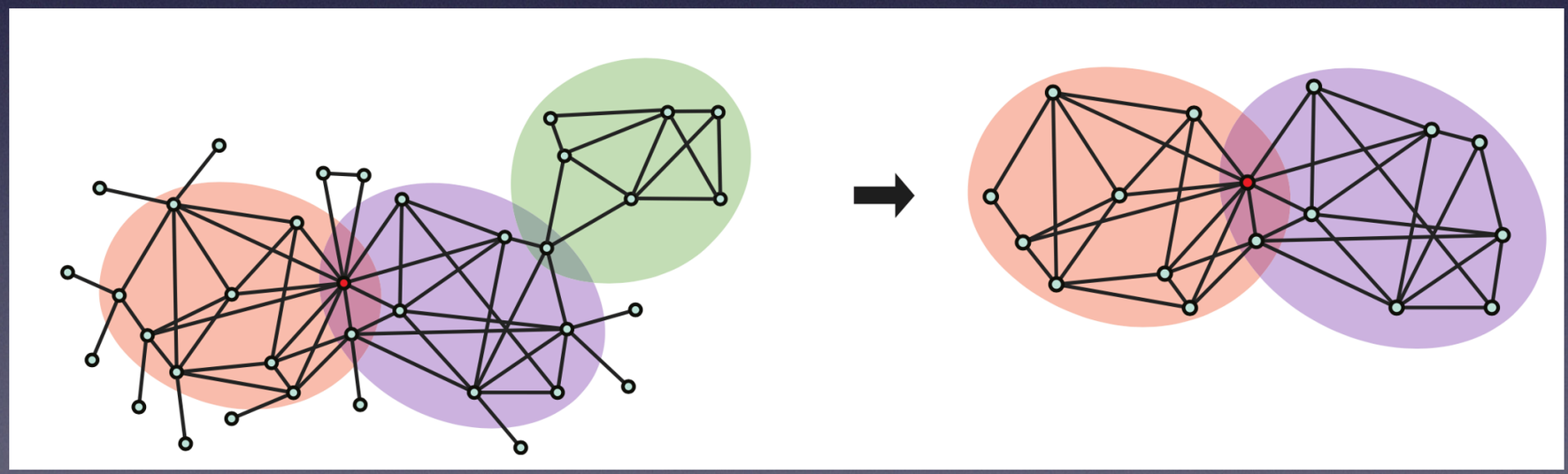
Evaluation on the community weight F

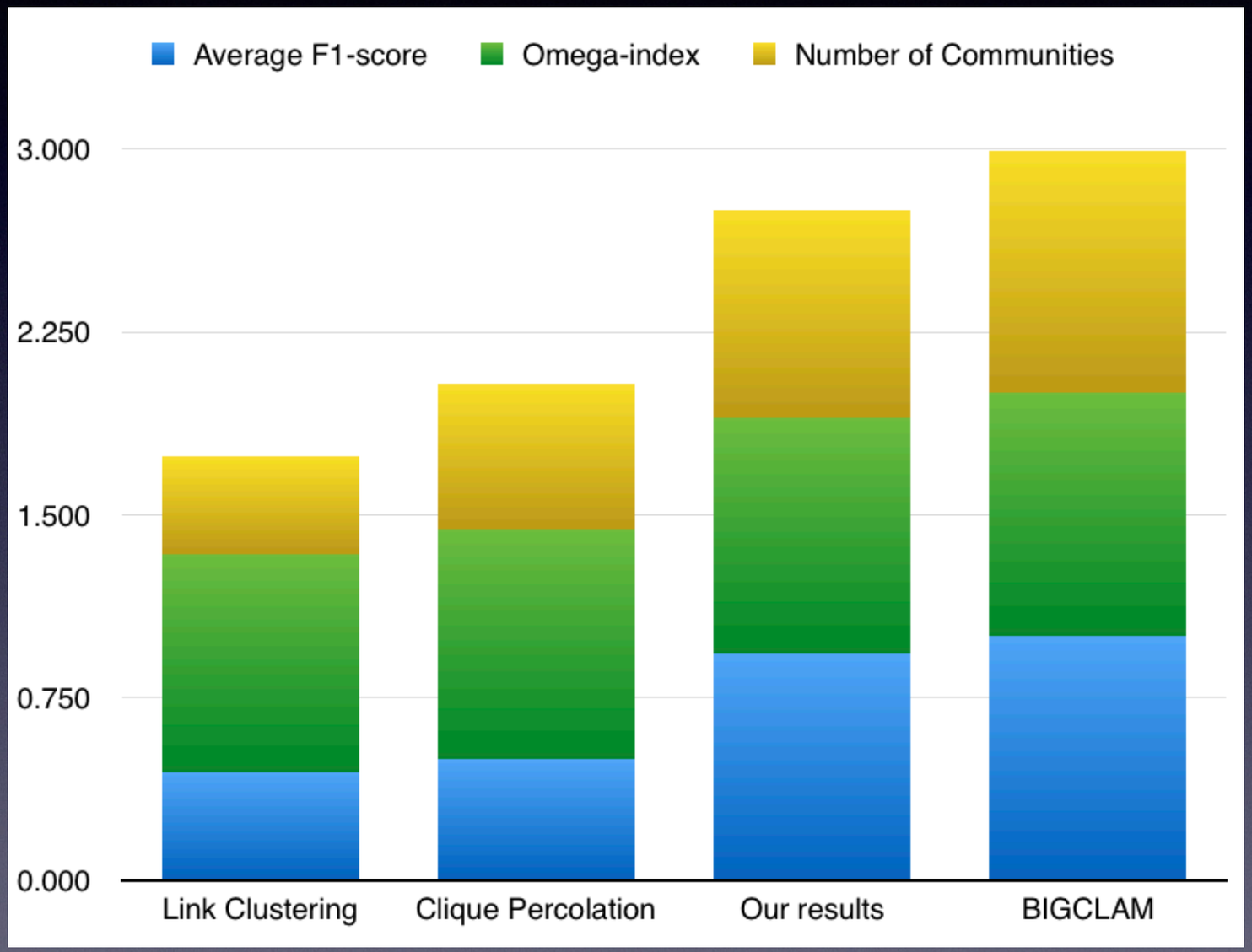
- Average F1 Score
- Omega index
- Accuracy in the number of communities

Some baseline methods do not scale well.

Solution: Sample subnetworks

- pick a random node u that belongs to at least two communities
- pick all the nodes that share at least one same community with u



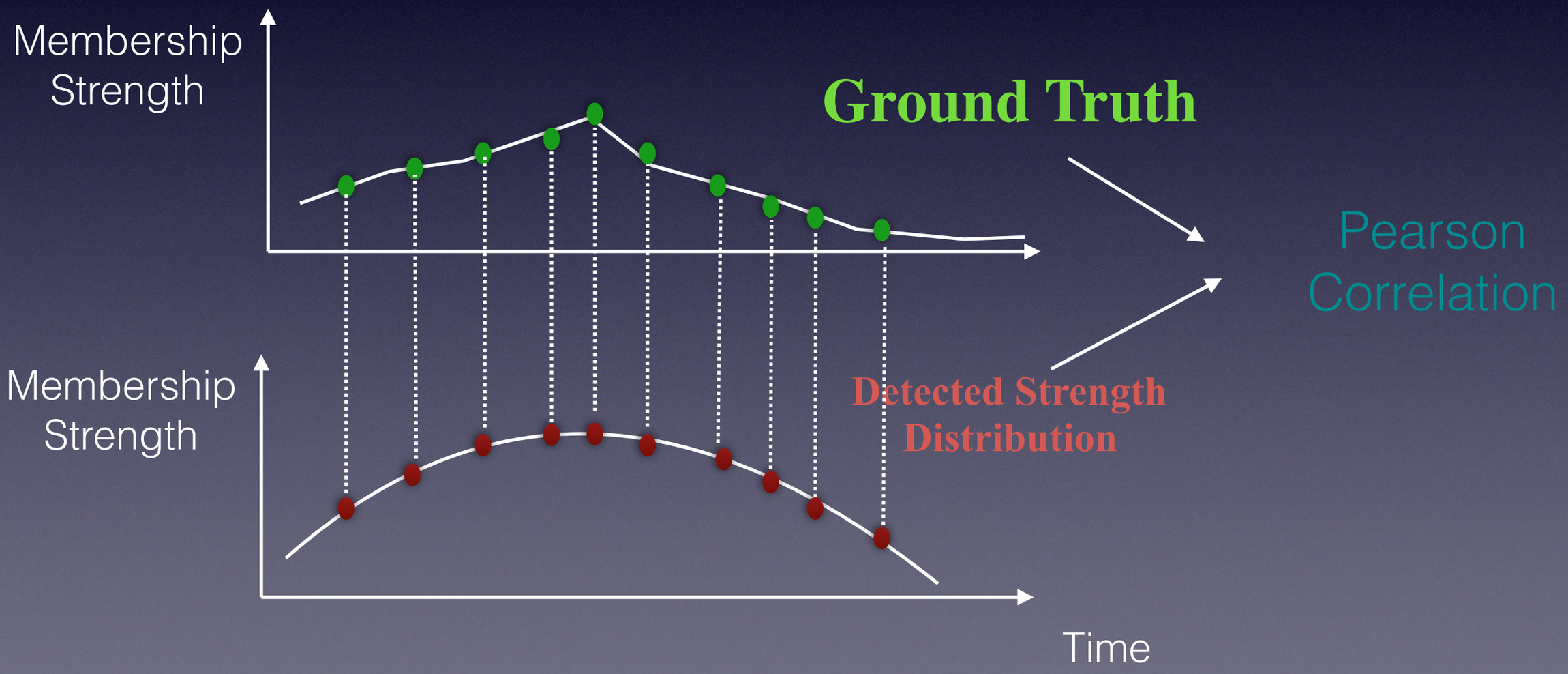

Average F1-score
Omega-index
Number of Communities
3.000
2.250
1.500
0.750
0.000
Link Clustering
Clique Percolation
Our results
BIGCLAM

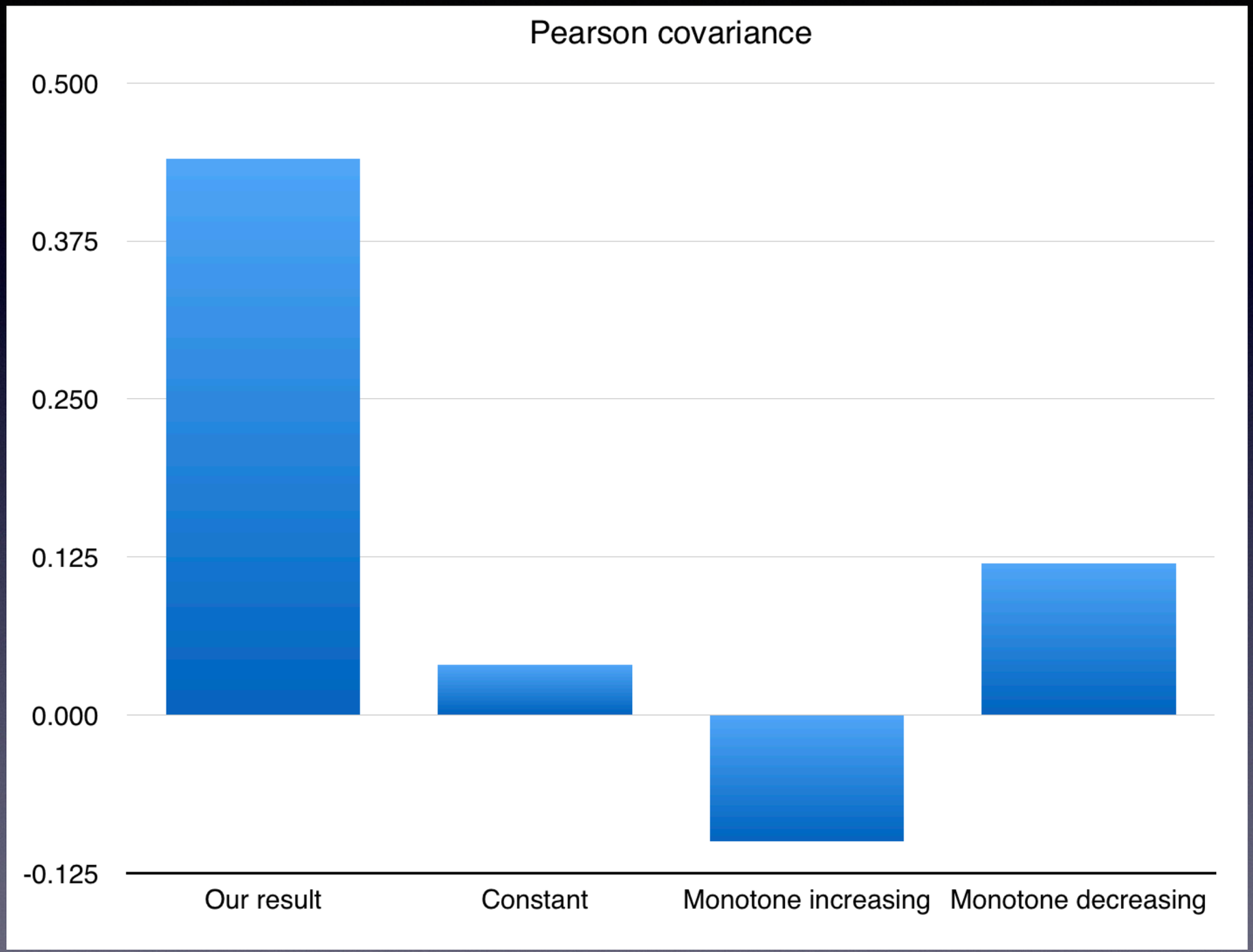
Two aspects

- The community weight (F)
- The temporal dimension (μ , σ)

Evaluation on the estimated temporal factors (μ , σ)

- Pearson Correlation




Pearson covariance
0.500
0.375
0.250
0.125
0.000
-0.125
Our result
Constant
Monotone increasing
Monotone decreasing

Challenges

- Dataset too large
- Fitting process very slow
- May suffer from local minimum

Future improvement

- Improve gradient ascent algorithm for faster speed
- Find better smaller datasets
- Use normalization or regularization for the parameters

Thank You!