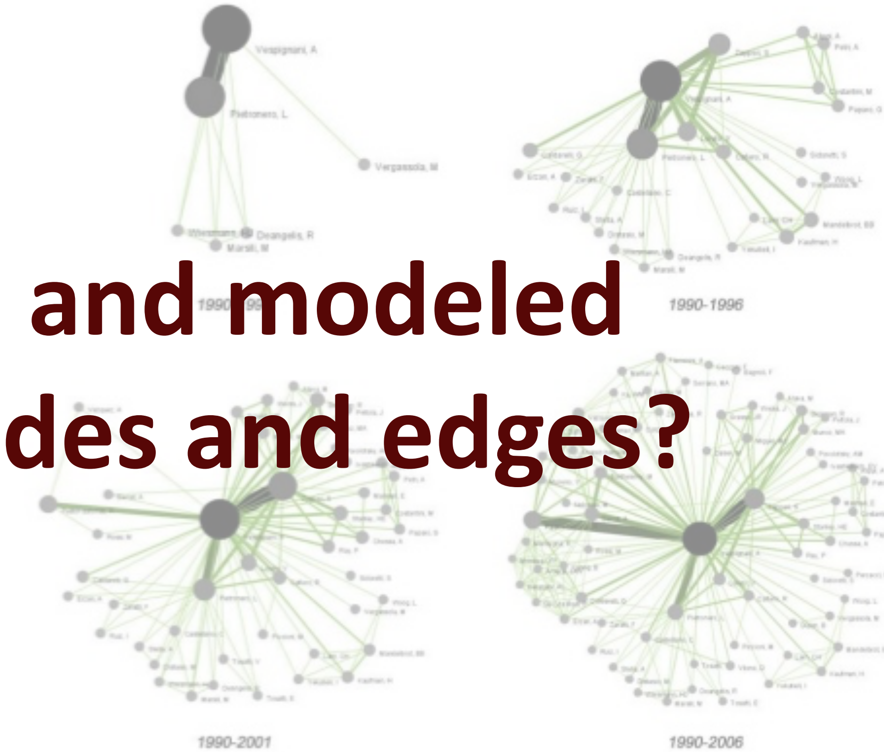
Dynamic Network Model from Partial Observations

Elahe Ghalebi [1], Baharan Mirzasoleiman [2], Radu Grosu [1], Jure Leskovec [2]

[1] TU Wien and [2] Stanford University

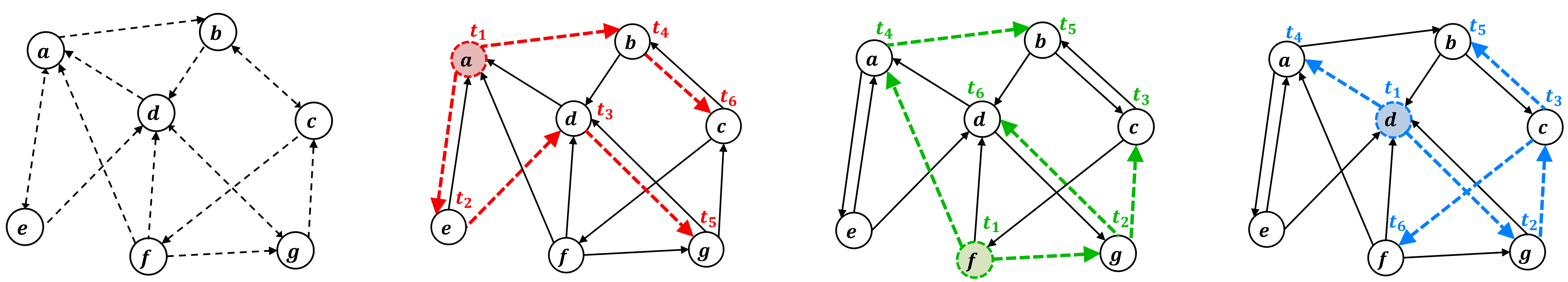
NeurIPS 2018

STANFORD



Can evolving network be inferred and modeled without directly observing their nodes and edges?

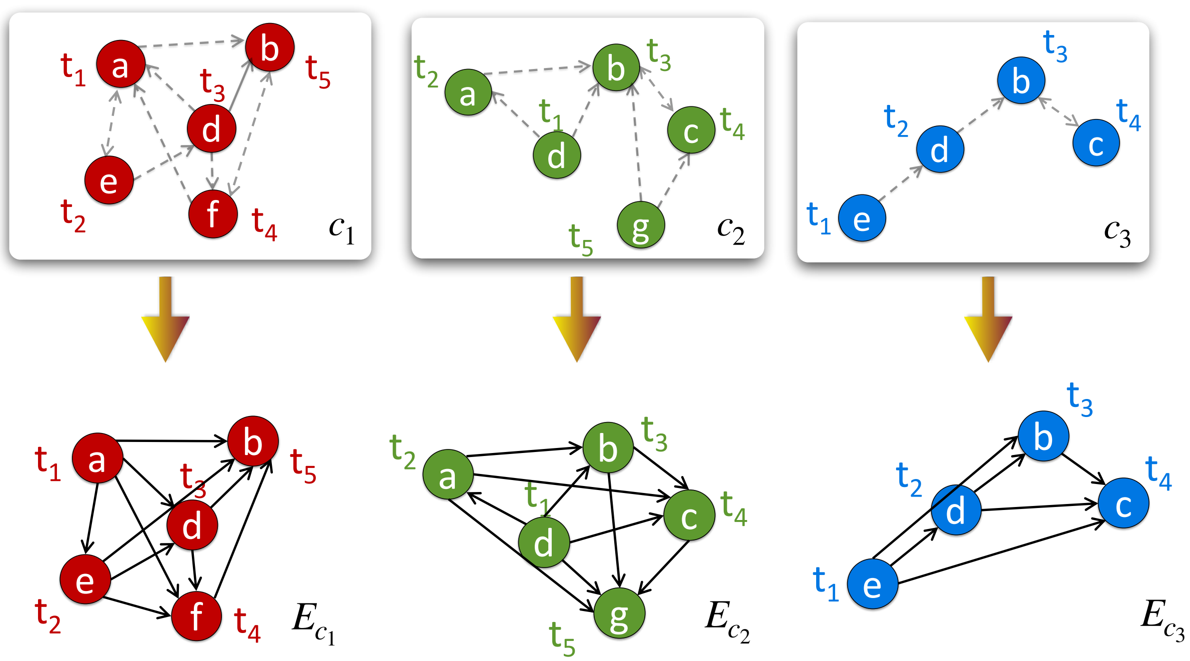
- In many applications, the edges of a dynamic network might not be observed
- We can only observe the dynamics of stochastic cascading process e.g. information diffusion, virus propagation occurring over the unobserved network



STANFORD

DYFERENCE framework

1- Extracting Observation from Diffusion Data



$t_1 < t_2 < \dots$

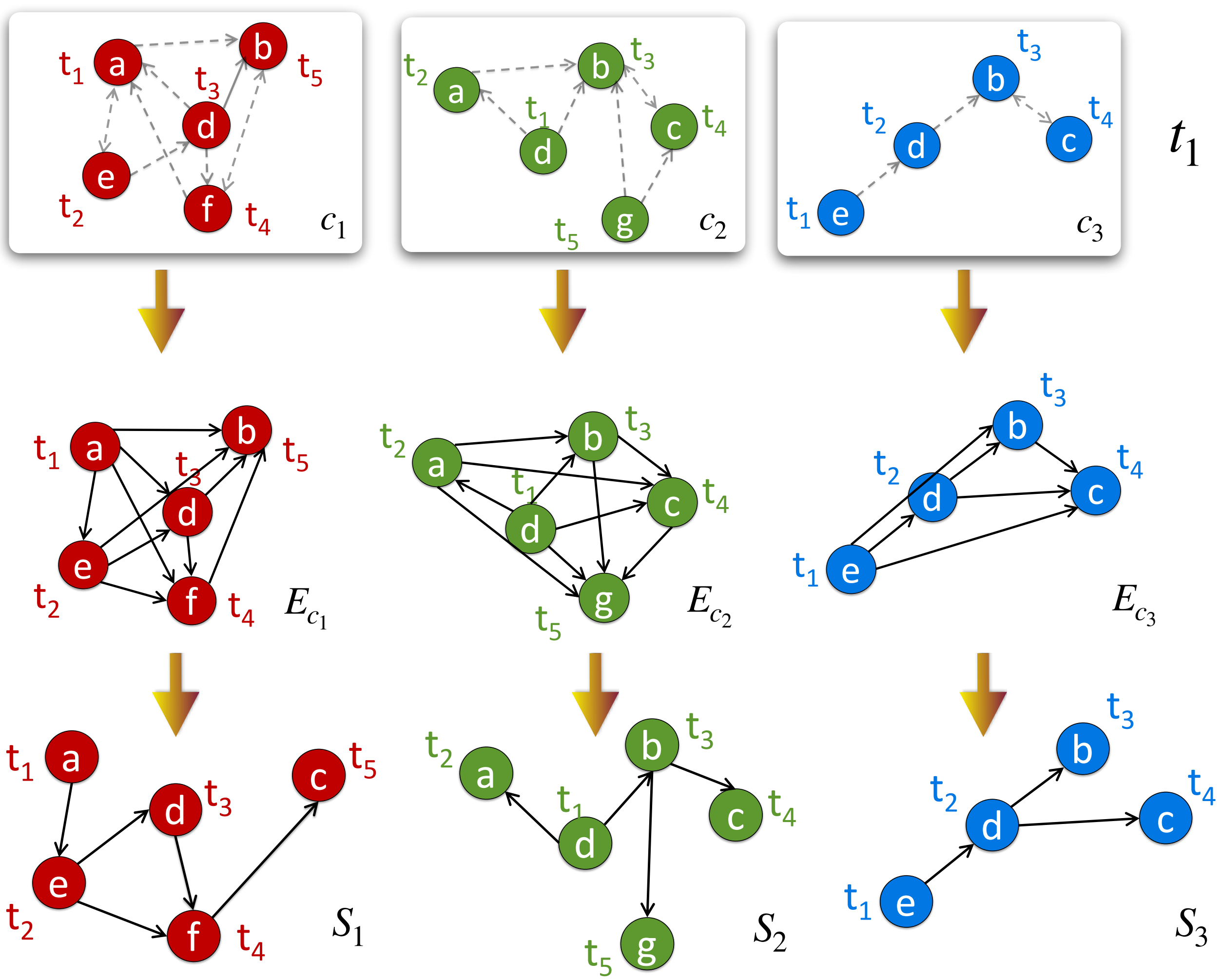
Find the set of possible edges in each cascade c_i as $E_{c_i} = \{ e_{uv} \,|\, t_u^{c_i} < t_v^{c_i} < \infty \}$

TU WIEN STANFORD

DYFERENCE framework

1- Extracting Observation from Diffusion Data

Round #1



$t_1 < t_2 < \ldots$

Calculate probability distribution over edges consistent with each cascade E_{c_i}

Calculate marginal probability of every edge in each E_{c_i}

Sample a set S_{c_i} of $\theta(|E_{c_i}|)$ edges based on marginal probabilities

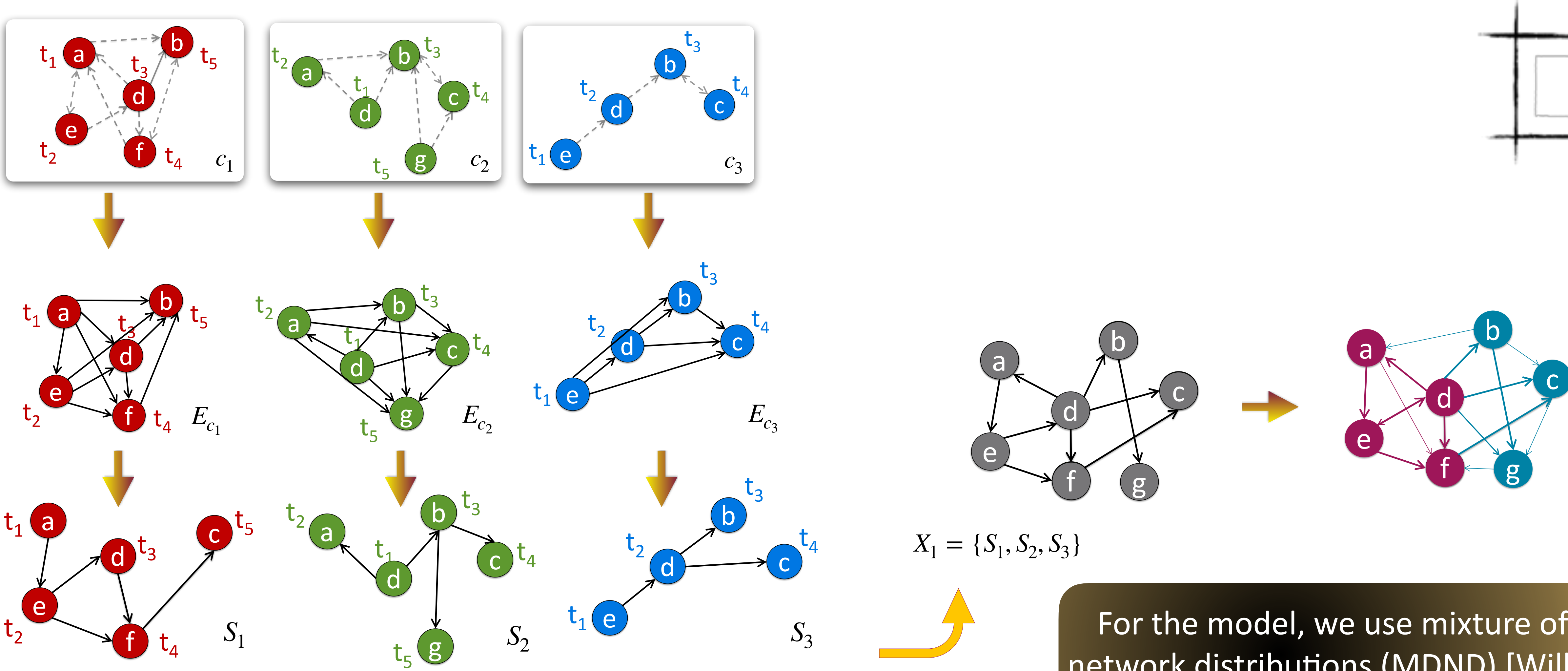
TU WIEN STANFORD

DYFERENCE framework

1- Extracting Observation from Diffusion Data

2- Update the model with the extracted observation X_1 using a collapsed Gibbs sampler

Round #1



For the model, we use mixture of Dirichlet network distributions (MDND) [Williamson'16]

Calculate probability distribution over edges consistent with each cascade E_{c_i}

Calculate marginal probability of every edge in each E_{c_i}

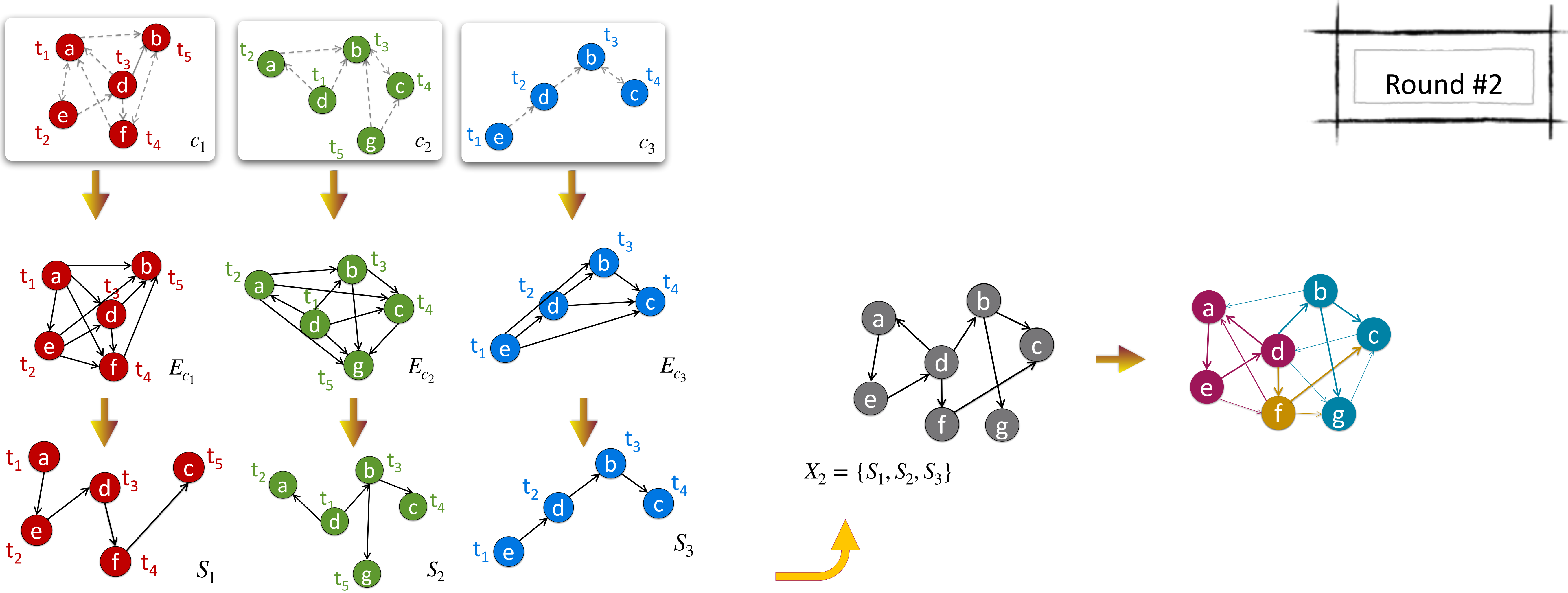
Sample a set S_{c_i} of $\theta(|E_{c_i}|)$ edges based on marginal probabilities

TU WIEN STANFORD

DYFERENCE framework

1- Extracting Observation from Diffusion Data

2- Update the model with the extracted observation X_2 using a collapsed Gibbs sampler



Calculate probability distribution over each E_{c_i} using updated edge probabilities from model

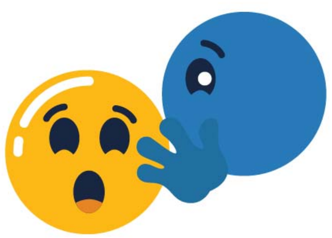
Calculate marginal probability of every edge in each E_{c_i}

Sample a set S_{c_i} of $\theta(|E_{c_i}|)$ edges based on marginal probabilities

TU WIEN STANFORD

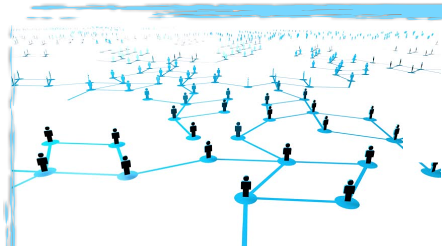
Online Dynamic Network Inference

1. Discretize time into intervals with length ω



2. Consider only infection times in current interval

$$t^c \in [(i-1)\omega, i\omega], \quad \forall c \in C$$



3. Update model with the observations in current ω

TU WIEN STANFORD

Performance Evaluation

Dynamic Bankruptcy Prediction

European country's financial transactions: 1,197,116 transactions;103,497 companies

	2012			2014			2016		
MAP@k	@10	@20	@30	@10	@20	@30	@10	@20	@30
INFOPATH	4.0	5.3	6.6	35.0	34.5	30.0	54.7	65.0	65.0
DYFERENCE	**17.6**	**19.1**	**20.6**	**62.0**	**51.9**	**38.1**	**69.6**	**85.7**	**85.7**
Hits@k	@10	@20	@30	@10	@20	@30	@10	@20	@30
INFOPATH	20.0	25.0	26.6	50.0	55.0	50.0	80.0	65.0	65.0
DYFERENCE	**40.0**	**45.0**	**46.6**	**70.0**	**65.0**	**50.0**	**80.0**	**70.0**	**70.0**

Our algorithm significantly outperforms the baselines

TU WIEN STANFORD

Conclusion

✓ Our algorithm provides a ***generative probabilistic model*** which:

- ✦ Identifies the underlying time-varying community structure
- ✦ Obtains dynamic predictive distribution over the edges
- ✦ Can be used for diffusion prediction, predicting the most influential nodes, and bankruptcy prediction

Poster: Today (Wed Dec 5th. @ Room 210 & 230) #7