ICWSM-2022
Atlanta, Georgia

Limits of Multilayer Diffusion Network Inference in Social Media Research

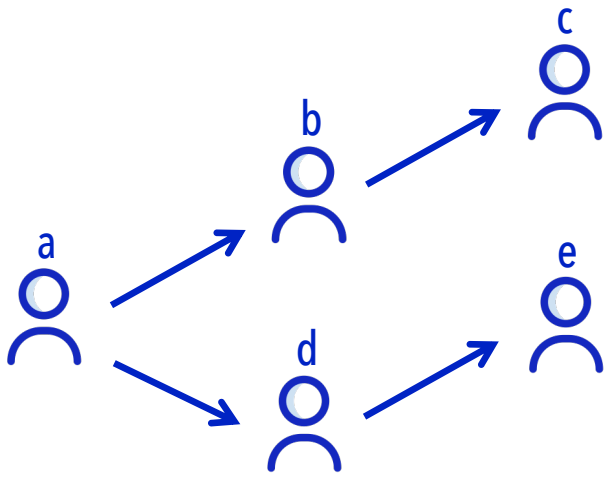
Yan Xia
Aalto University

Ted Hsuan Yun Chen
Aalto University
University of Helsinki

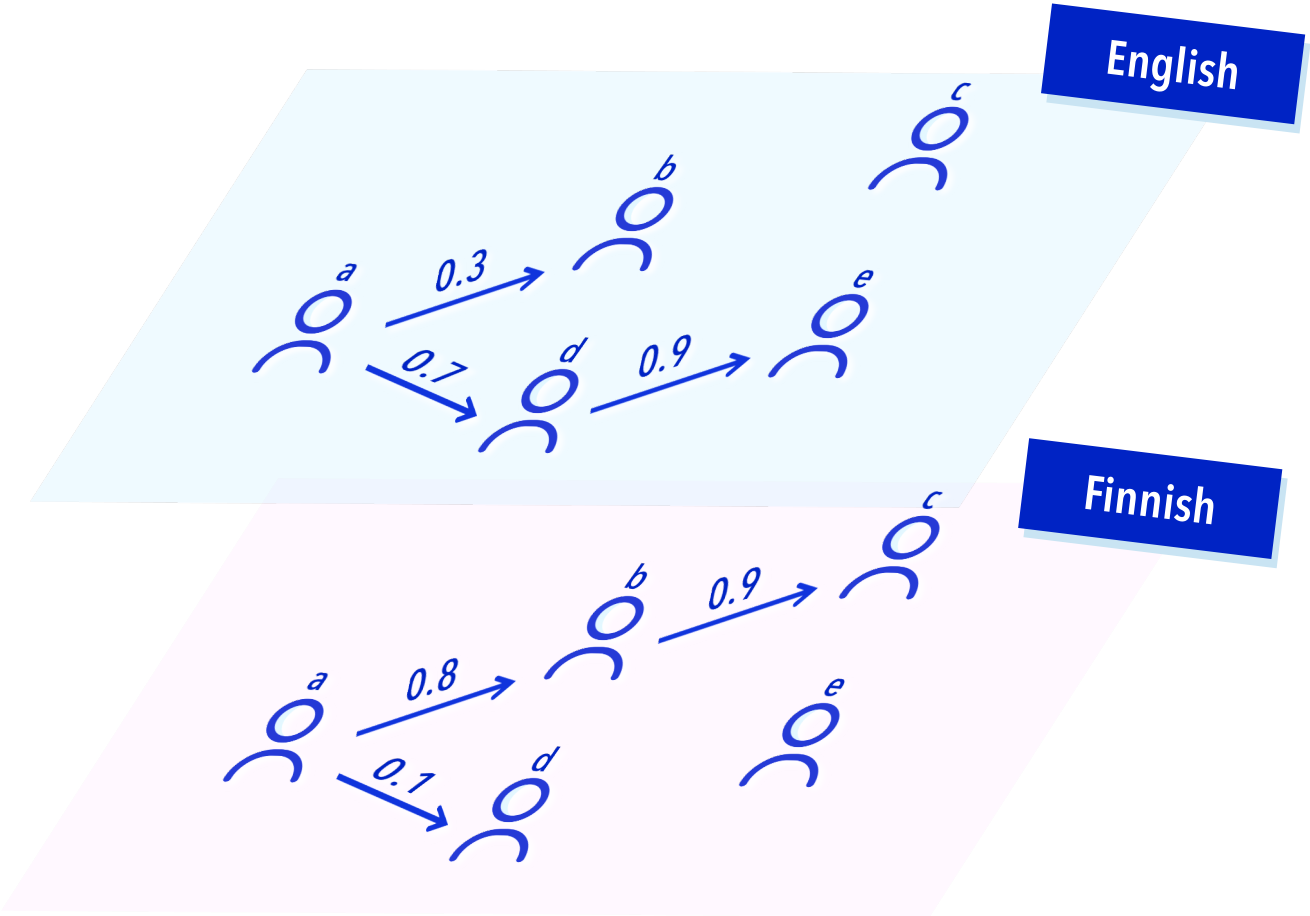
Mikko Kivelä
Aalto University

Aalto University

Motivation: Multilayer diffusion network

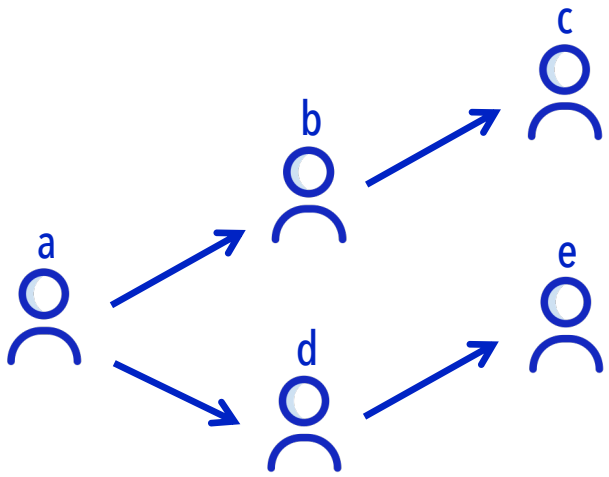


Single layer follower network

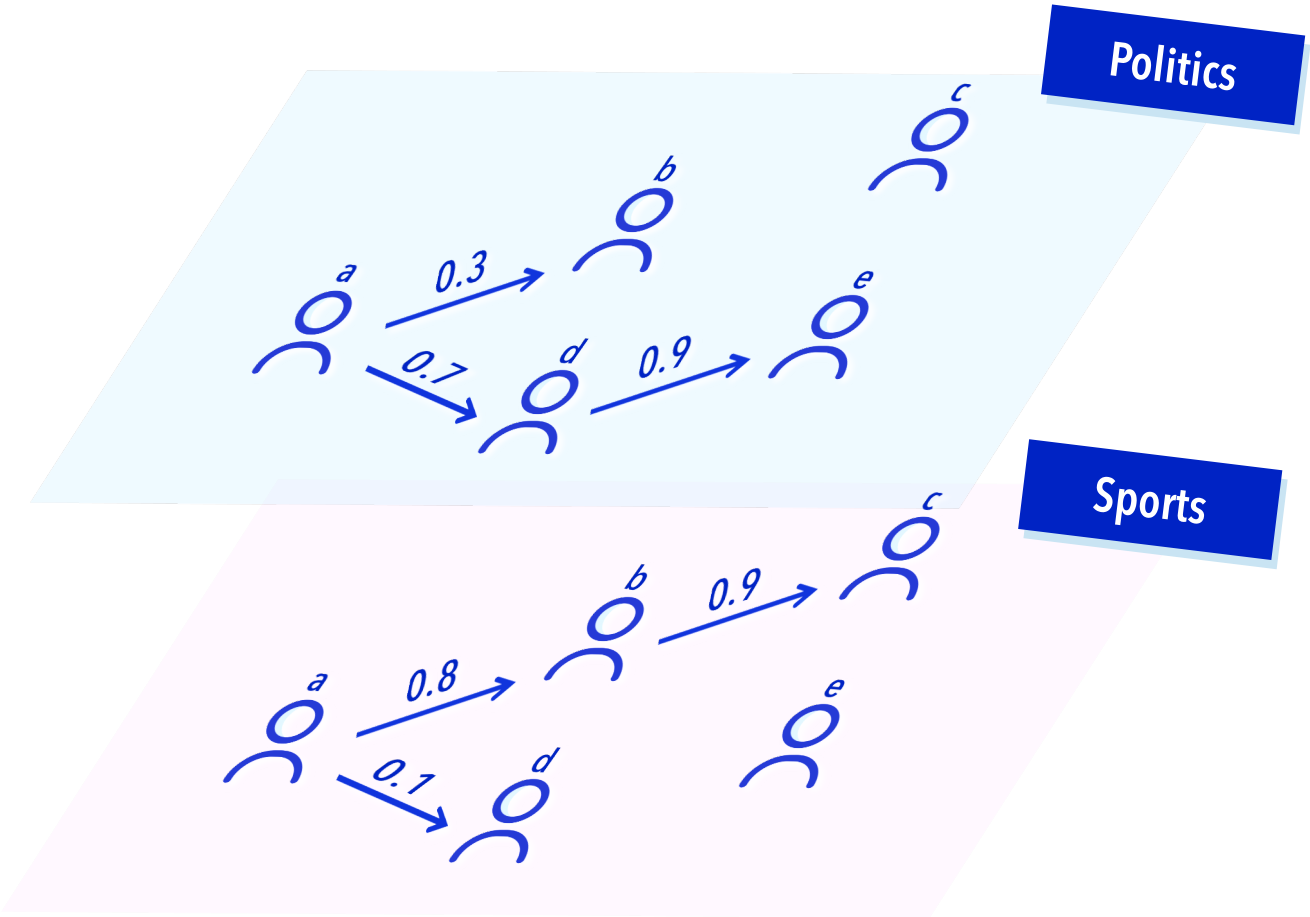


Multilayer diffusion network

Motivation: Multilayer diffusion network

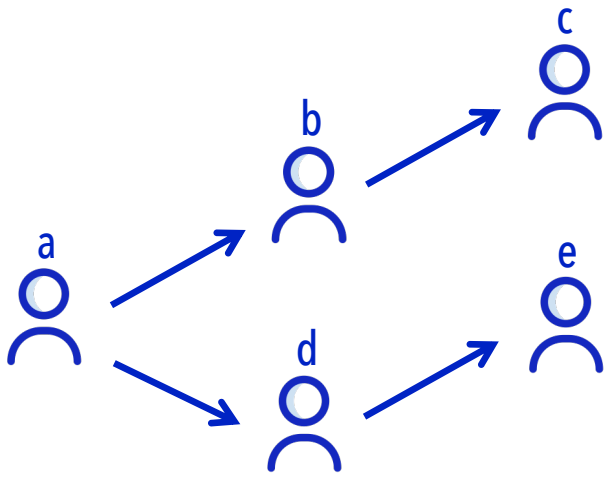


Single layer follower network

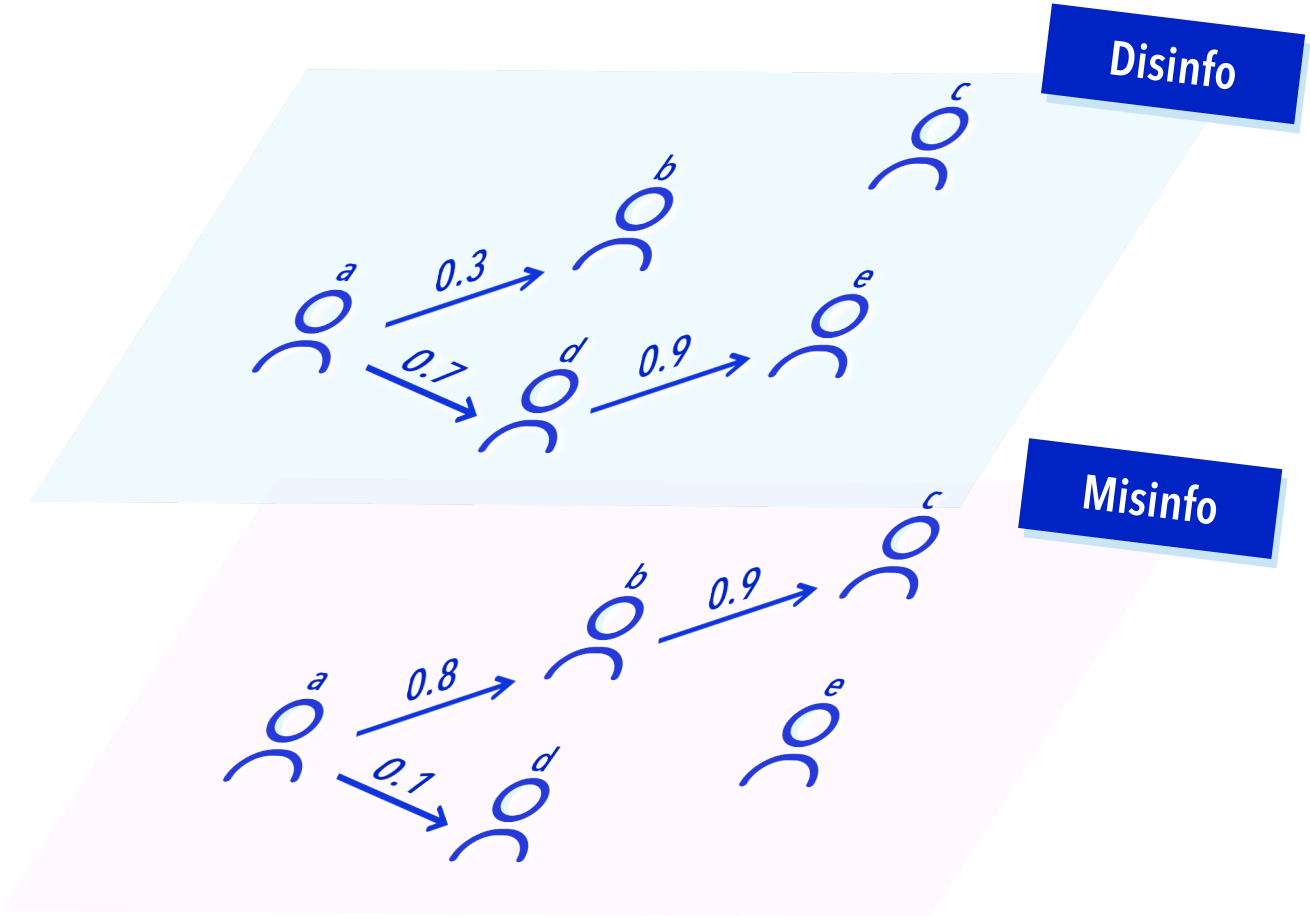


Multilayer diffusion network

Motivation: Multilayer diffusion network

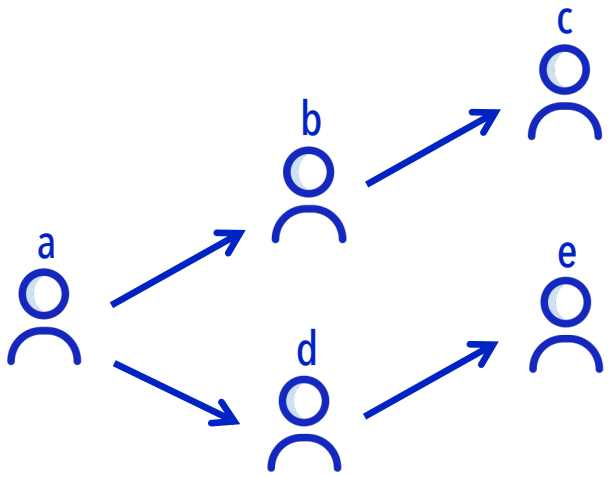


Single layer follower network

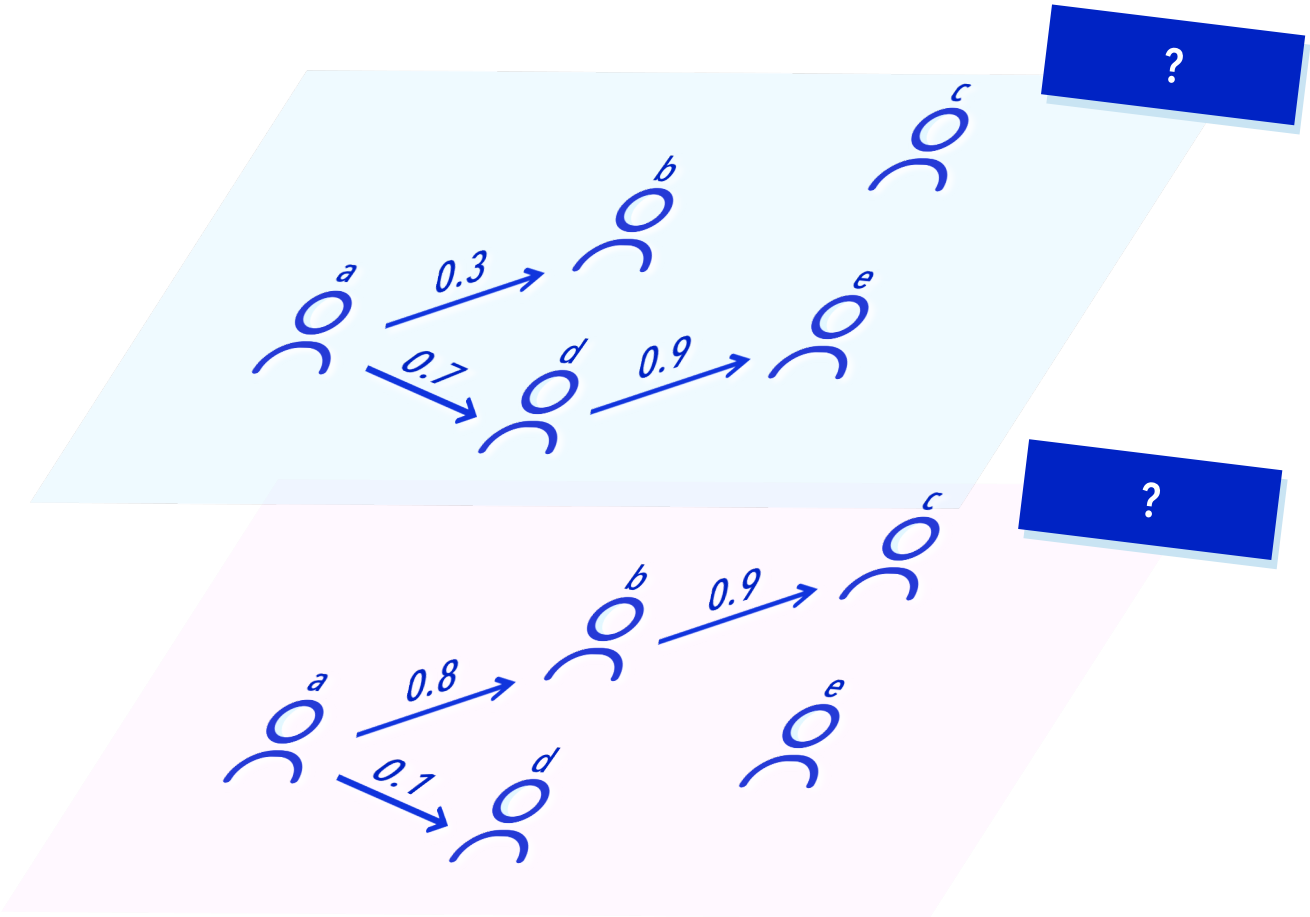


Multilayer diffusion network

Motivation: Multilayer diffusion network

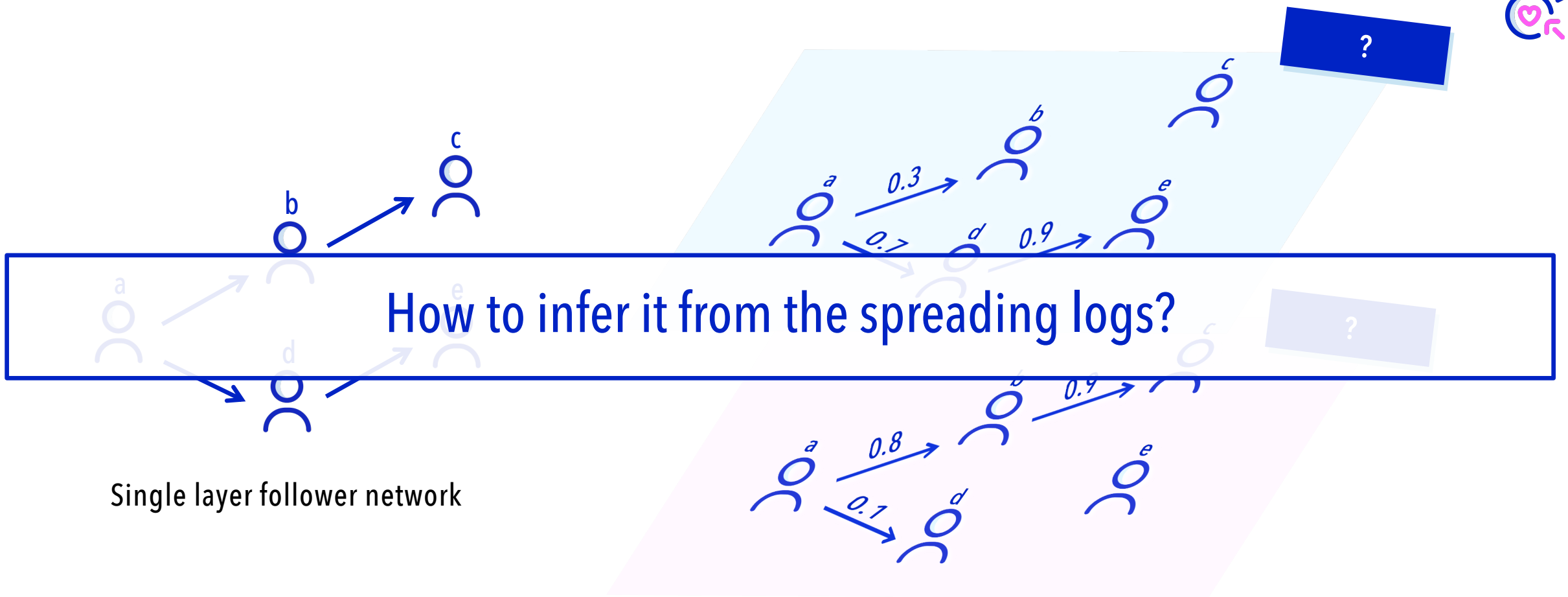


Single layer follower network



Multilayer diffusion network

Motivation: Multilayer diffusion network



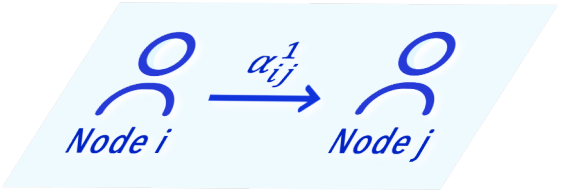
Generative spreading model [1]

Layer-wise edge weights α

Cascade membership π

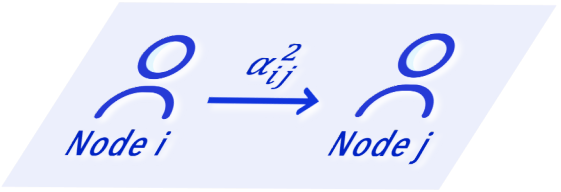
Cascade c

Layer 1



π_1^c

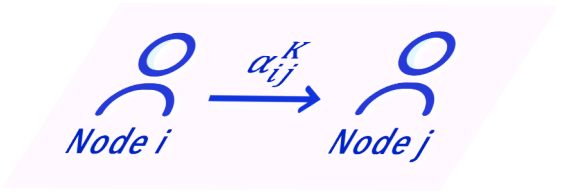
Layer 2



π_2^c

⋮ ⋮

Layer K



π_K^c

$\sum_{k=1}^{K} \pi_k^c = 1$

Transmission rate of cascade c from i to j

$$\lambda_{ij}^c = \sum_{k=1}^{K} \pi_k^c \alpha_{ij}^k$$

Transmission time of cascade c from i to j

$$\Delta t_{ij}^c \sim Exp(\lambda_{ij}^c)$$

Probability that i succeeds/fails to activate j on c

$$p_{ij}^c = f(\Delta t_{ij}^c; \lambda_{ij}^c) \qquad f(t) = P(T = t)$$

$$\overline{p_{ij}^c} = S(T - t_i^c; \lambda_{ij}^c) \qquad S(t) = P(T \geq t)$$

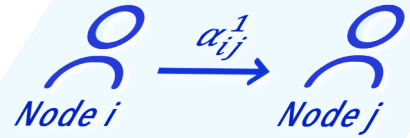
[1] Wang, S., Hu, X., Yu, P. S., & Li, Z. (2014, August). MMRate: inferring multi-aspect diffusion networks with multi-pattern cascades. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1246-1255).

Generative spreading model [1]

Layer-wise edge weights α

Cascade membership π

Layer 1



Cascade c

π_1^c

Transmission rate of cascade c from i to j

$$\lambda_{ij}^c = \sum_{k=1}^{K} \pi_k^c \alpha_{ij}^k$$

Minimize

$$-\log \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\alpha}) = \sum_c \sum_{j:t_j^c<T} \Big(\sum_{u:t_u^c<t_j^c} \Delta t_{uj}^c \sum_{k=1}^{K} \pi_k^c \alpha_{uj}^k - \log \sum_{i:t_i^c<t_j^c} \sum_{k=1}^{K} \pi_k^c \alpha_{ij}^k + \sum_{n:t_n^c>T} (T - t_n^c) \sum_{k=1}^{K} \pi_k^c \alpha_{jn}^k \Big)$$

Layer K

π_K^c

$$\sum_{k=1}^{K} \pi_k^c = 1$$

Probability that i succeeds/fails to activate j on c

$$p_{ij}^c = f(\Delta t_{ij}^c; \lambda_{ij}^c) \qquad f(t) = P(T = t)$$

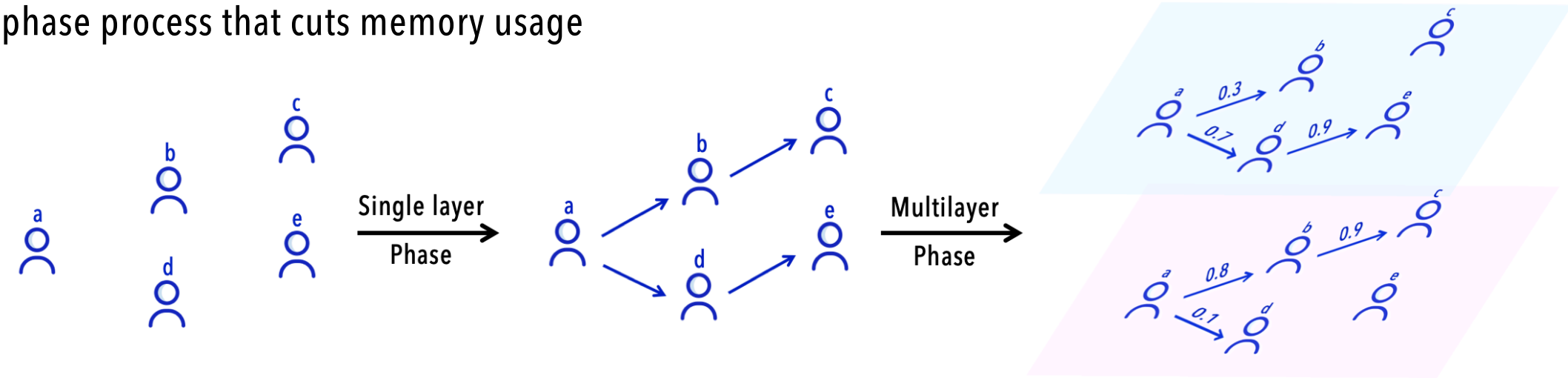
$$\overline{p_{ij}^c} = S(T - t_i^c; \lambda_{ij}^c) \qquad S(t) = P(T \geq t)$$

[1] Wang, S., Hu, X., Yu, P. S., & Li, Z. (2014, August). MMRate: inferring multi-aspect diffusion networks with multi-pattern cascades. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1246-1255).

Implementation

Uses GPU computing for accelerating the optimization

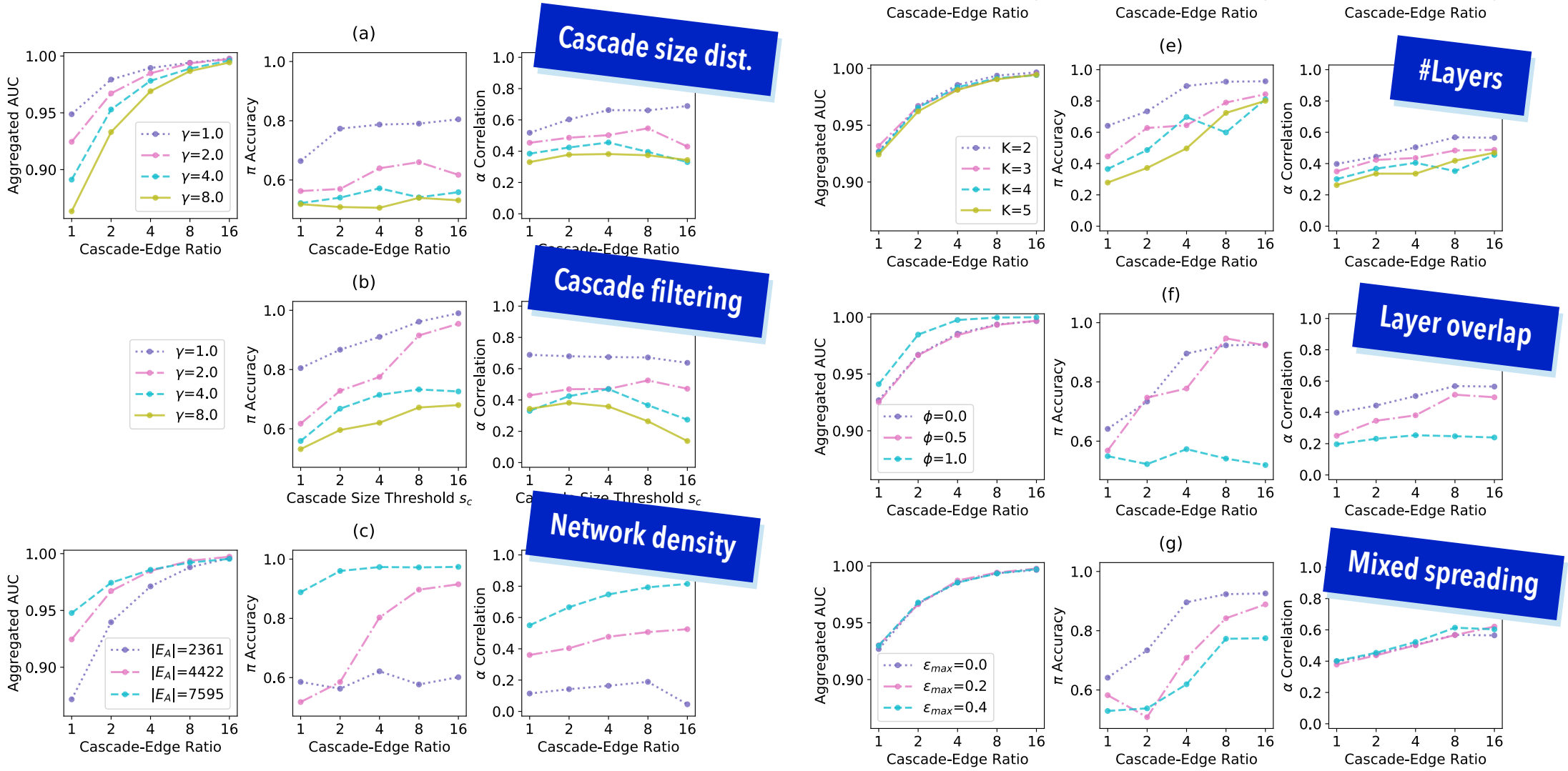
Two-phase process that cuts memory usage



Outperforms existing implementations in accuracy within comparable runtime

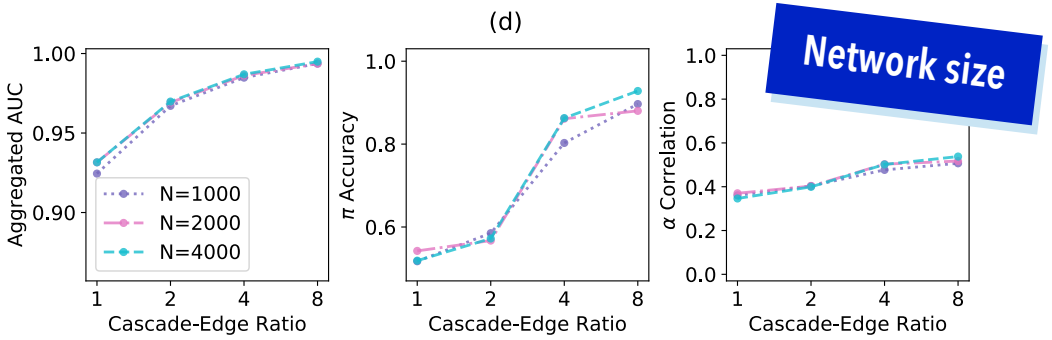
But realistic!

Synthetic data testing: Inference accuracy under varied …



But realistic!

Synthetic data testing: Inference accuracy under varied …



Takeaway 1: Excluding small cascades increases multilayer inference accuracy

Takeaway 2: The multilayer inference is still extremely inaccurate when
i) the proportion of large cascades is low OR
ii) the ground-truth network is sparse

Therefore:
i) We need to estimate the inference accuracy before applying the method to real data
ii) We need to develop a more real-data-applicable inference method

Thanks for listening!

Any question or comment?

Big thanks to: Aalto Scientific Computing and Barbara Keller

Code: https://github.com/ECANET-research/multic

Contact: **Yan Xia**

yan.xia@aalto.fi

@yanxxia