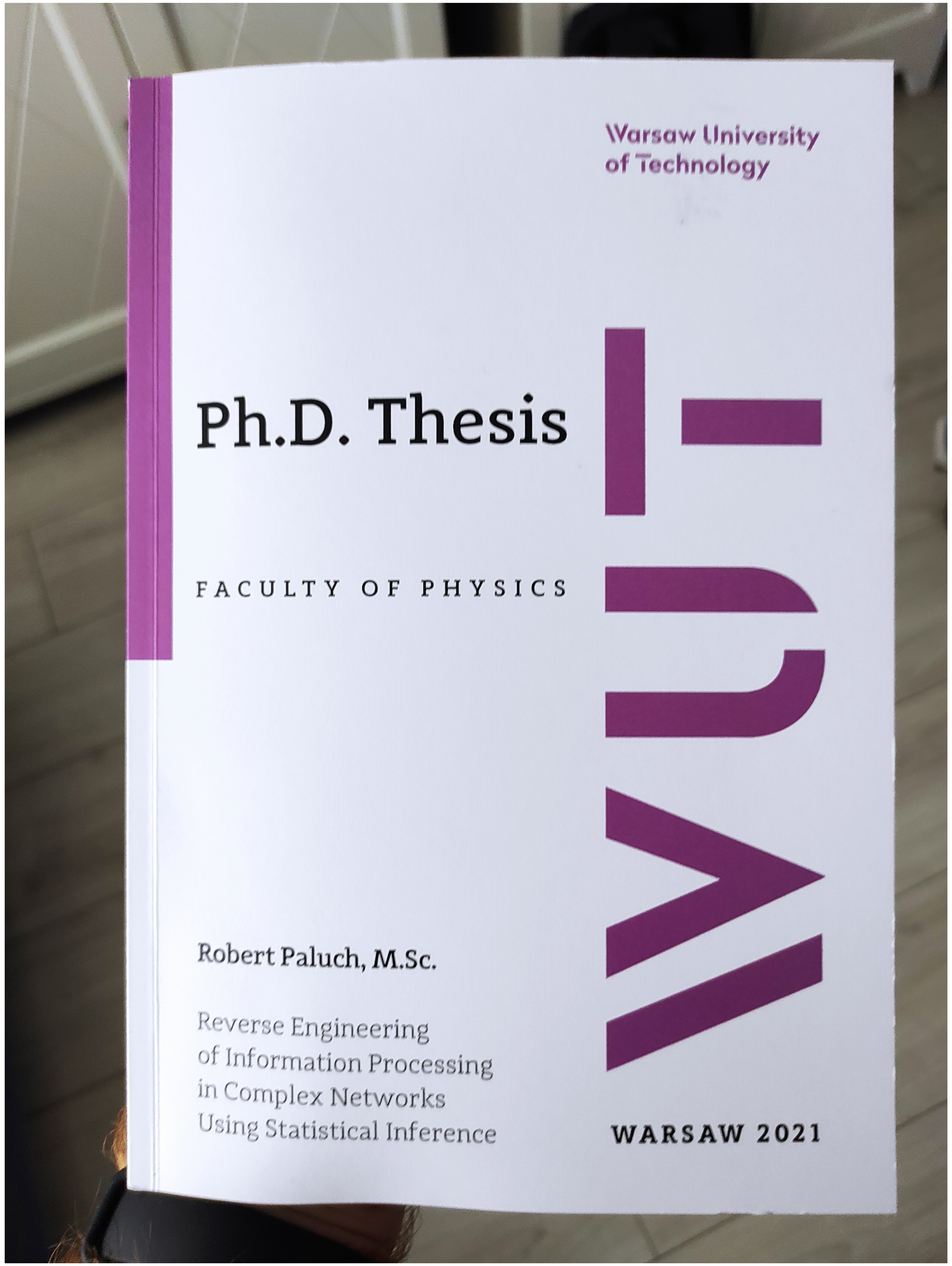
Wydział Fizyki

POLITECHNIKA WARSZAWSKA

Impact of multilayer topology on source localization in complex networks

Robert Paluch, Łukasz G. Gajewski, Krzysztof Suchecki, Janusz A. Hołyst

Politechnika Warszawska



Politechnika
Warszawska

Problem

The spreading of fake news, misinformation, pseudo-scientific views is a real threat to our society.

Politechnika
Warszawska

Source: Google Graphics

Ways to mitigate

8 useful steps for fighting fake news recommended by Internation Federation of Library Associtations and Institutions.

Politechnika Warszawska

Ways to mitigate

8 useful steps for fighting fake news recommended by Internation Federation of Library Associtations and Institutions.

The first author (or the source) is not always visible.

Politechnika Warszawska

Source localization methods

In the last decade, there were many papers considering the rumor source identification in complex networks.

Three main approaches can be distinguish:

- complete observation (we see states of all nodes in one time moment),
- snapshot (we see states of some subset of nodes in one time moment),
- detector-based (we have the infection times of some small part of nodes).

Source localization methods

Many variants of the problem:

- identification of the multiple sources,
- different spreading mechanism,
- time-varying topology of network.

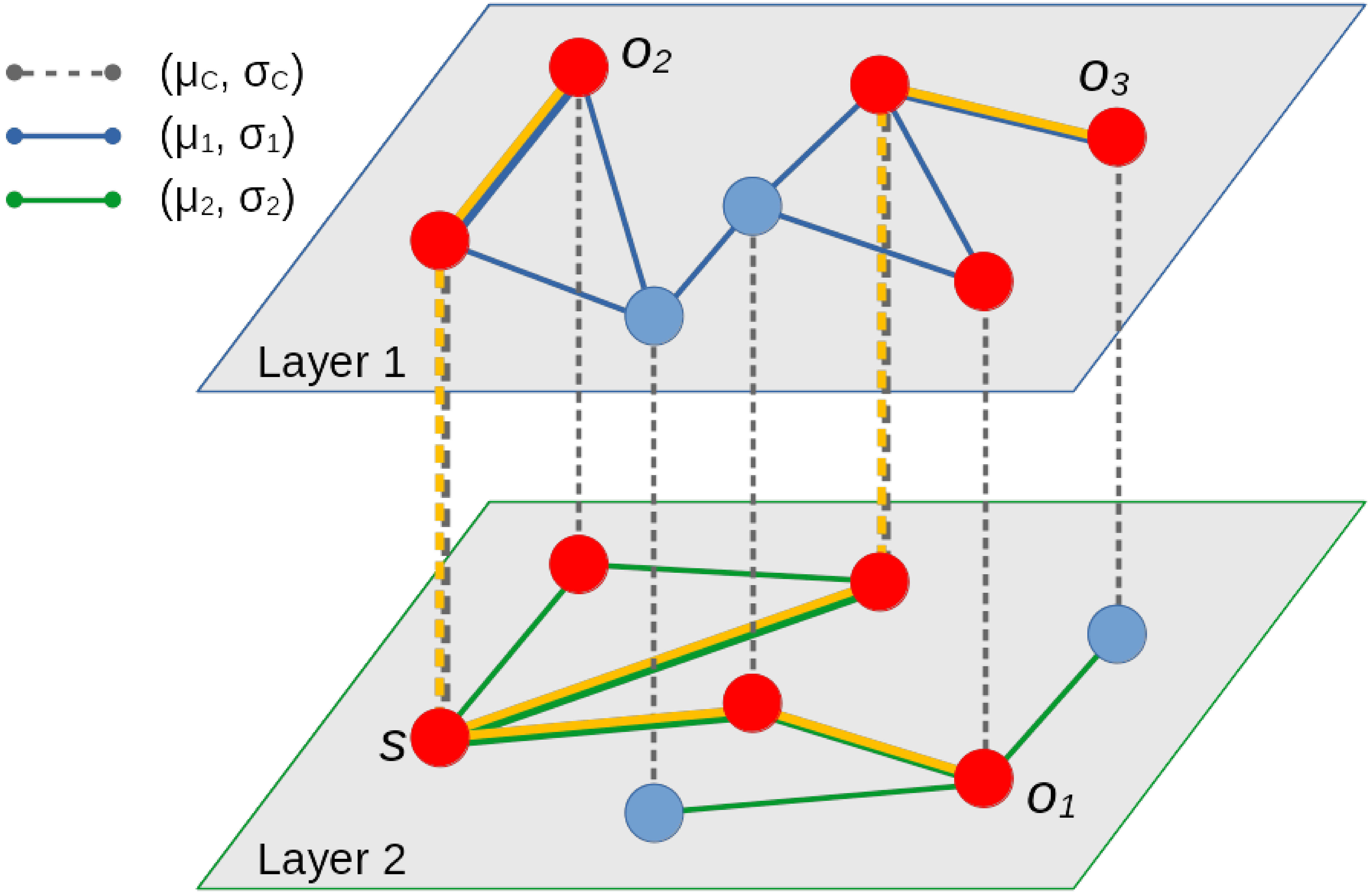
Source localization methods

Many variants of the problem:

- identification of the multiple sources,
- different spreading mechanism,
- time-varying topology of network.

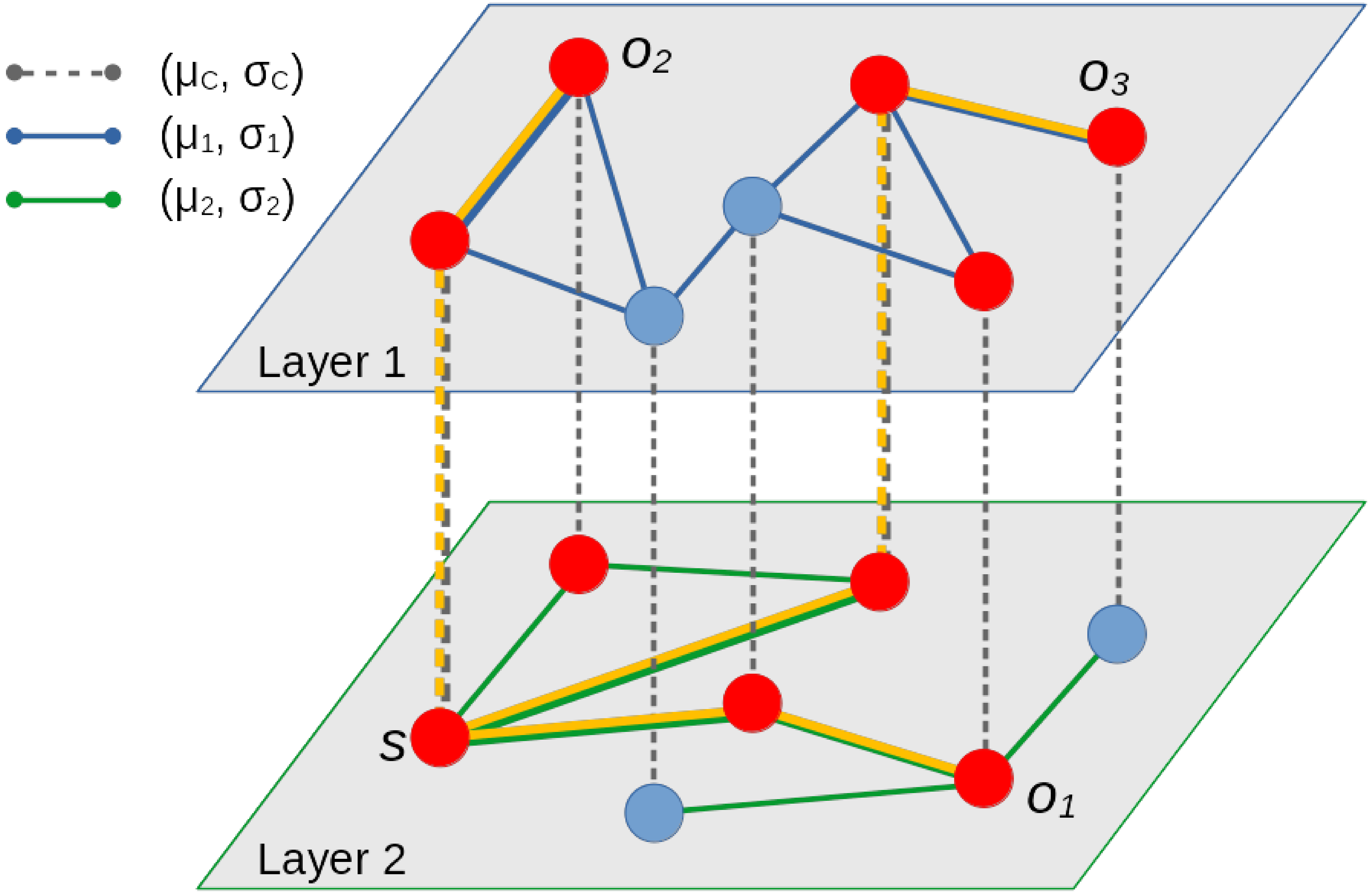
But what with the **multilayer structure**?
People operate in many social networks, which can have different properties, particularly **different speeds of spreading information**.

Information propagation modeling



Politechnika Warszawska

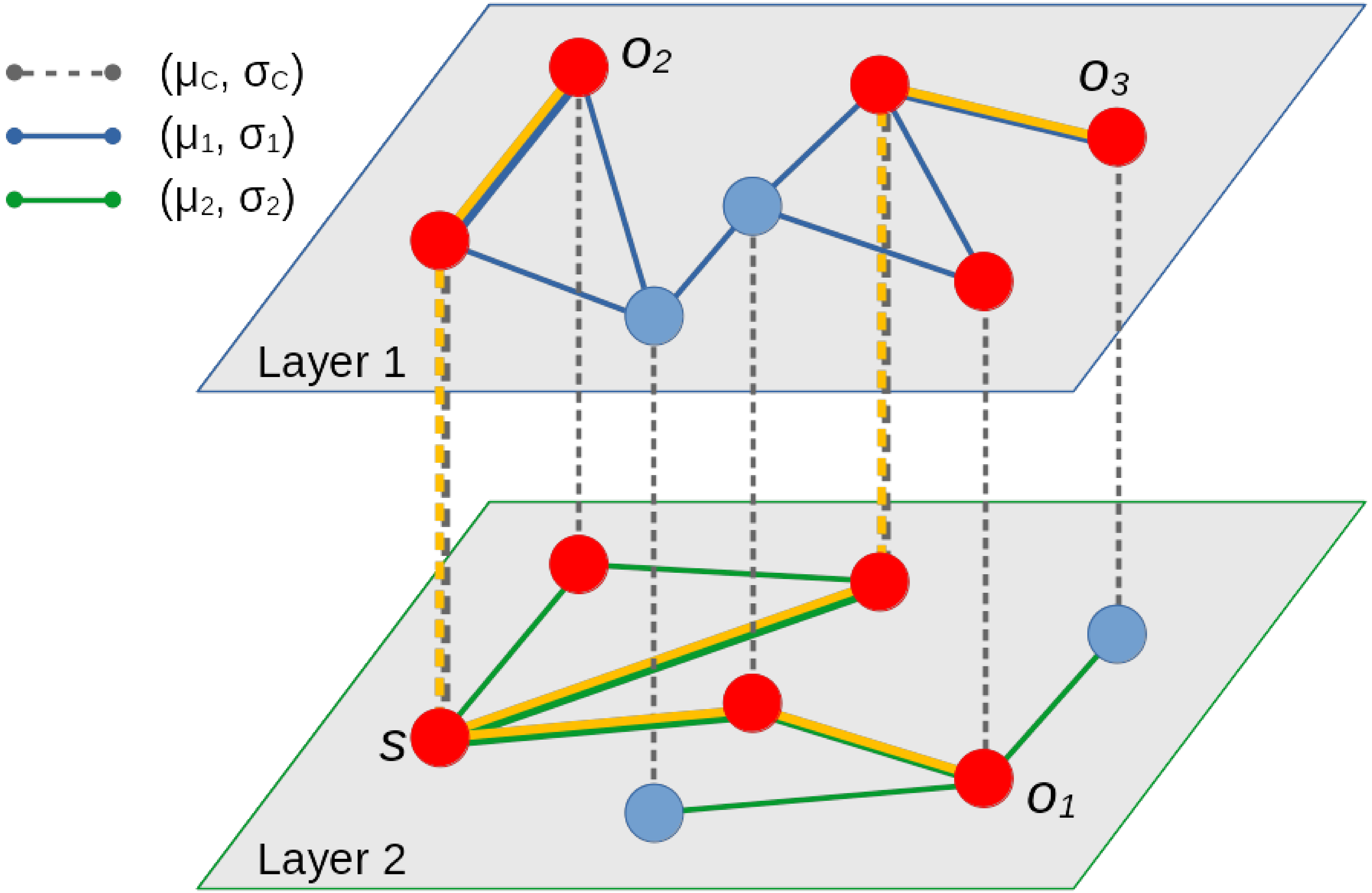
Information propagation modeling



- **The layers are interconnected (each with each).**
- No intermediate layers.
- The interlinks exists only between the replicas of the same user.

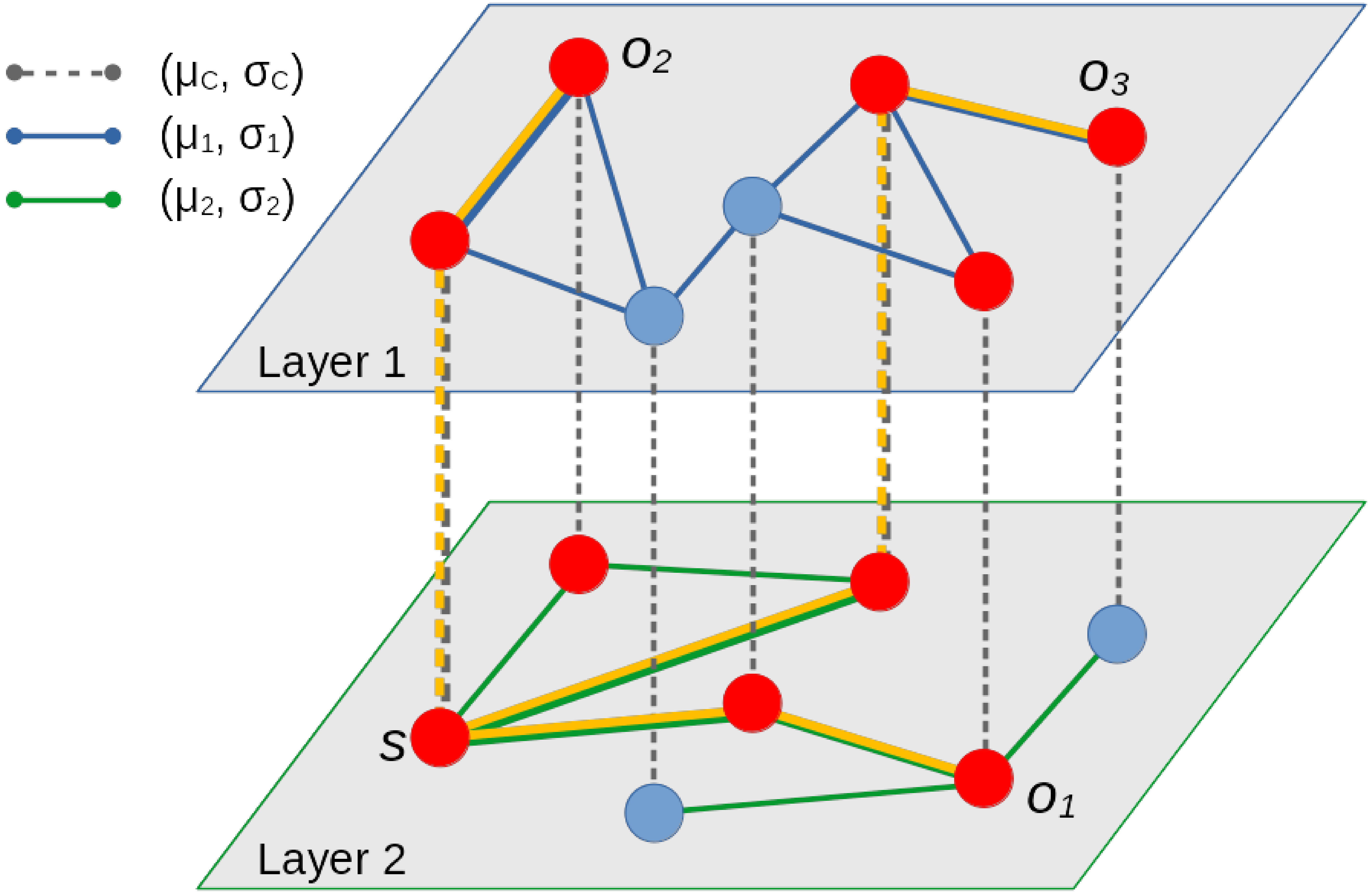
Information propagation modeling

- **Each layer can have different parameters of propagation.**
- Propagation process is described by the mean μ and the standard deviation σ of the time delays on links.
- Interlinks have their own μ and σ.



Information propagation modeling

- **The replicas of an agent can have different states in different layers.**
- People share the information obtained in the other layer with a delay.
- This delay plays a role of the coupling strength between layers.

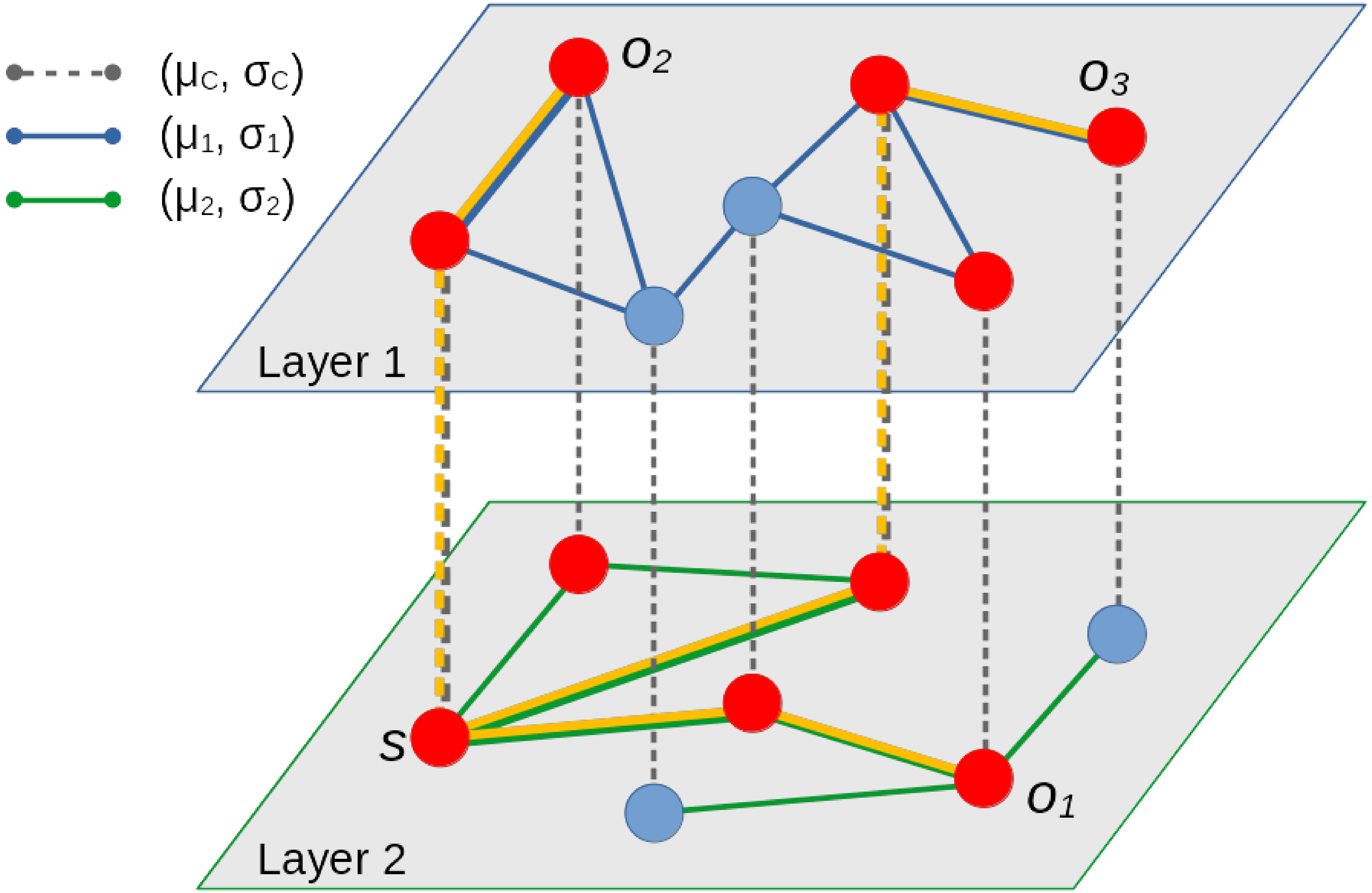


Politechnika Warszawska

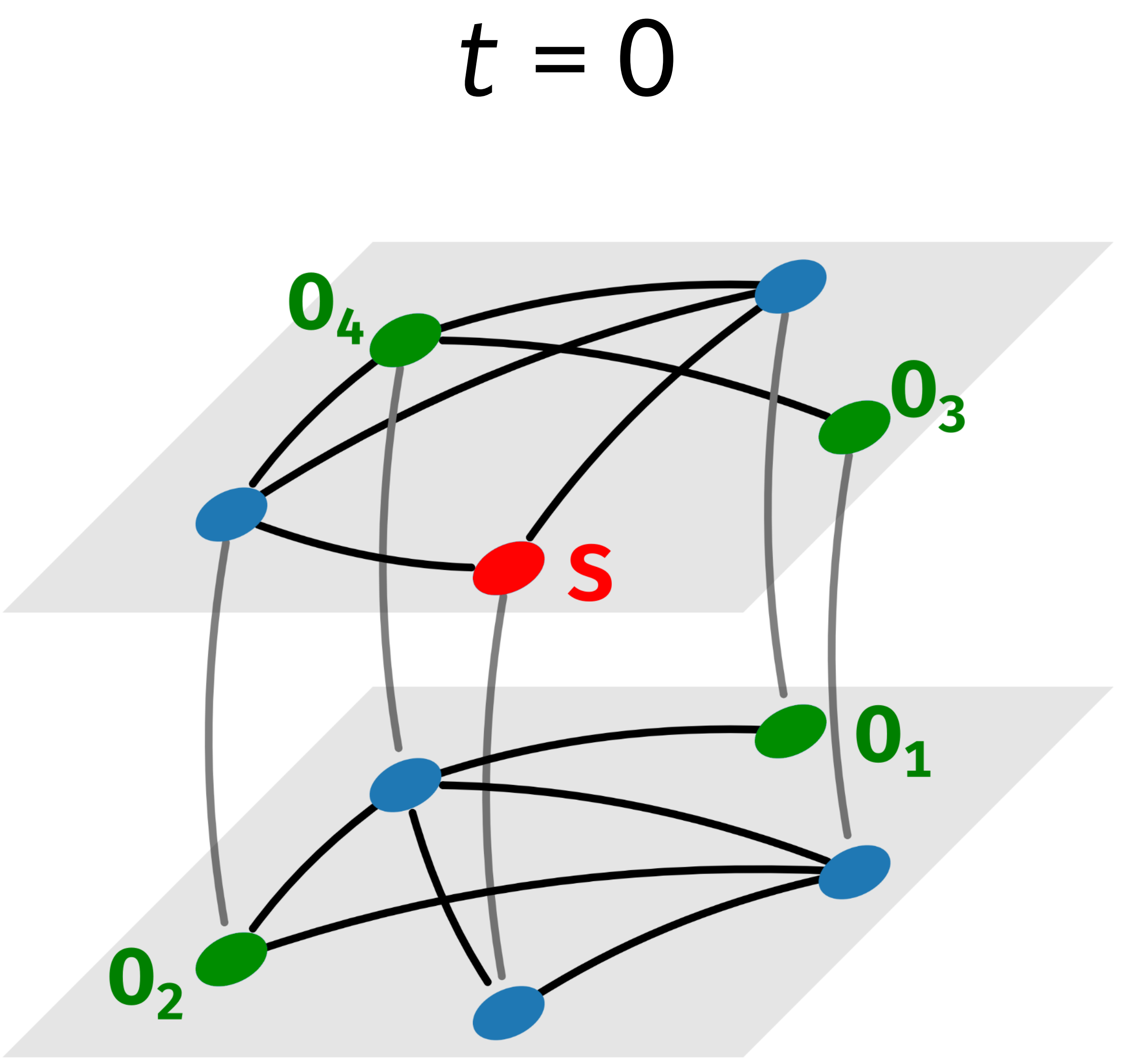
Information propagation modeling

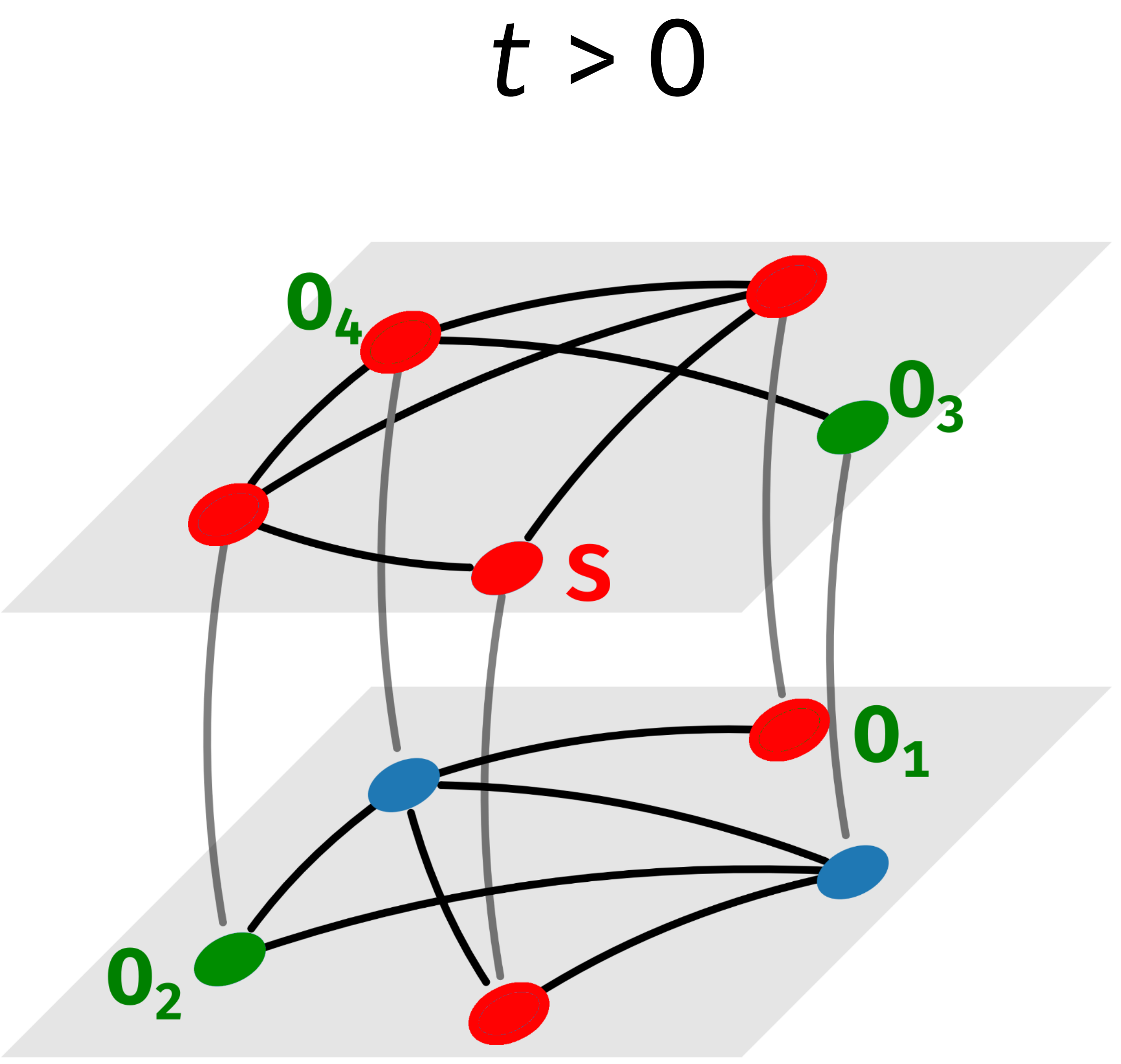
- **Information originates from one replica (single-source problem).**
- We use the Susceptible-Infected model to propagate the signal, which reduces the number of spreading parameters to only one – infection rate β.

$$\mu = \frac{1}{\beta} \qquad \sigma = \frac{\sqrt{1-\beta}}{\beta}$$



Observers-based maximum likelihood estimation



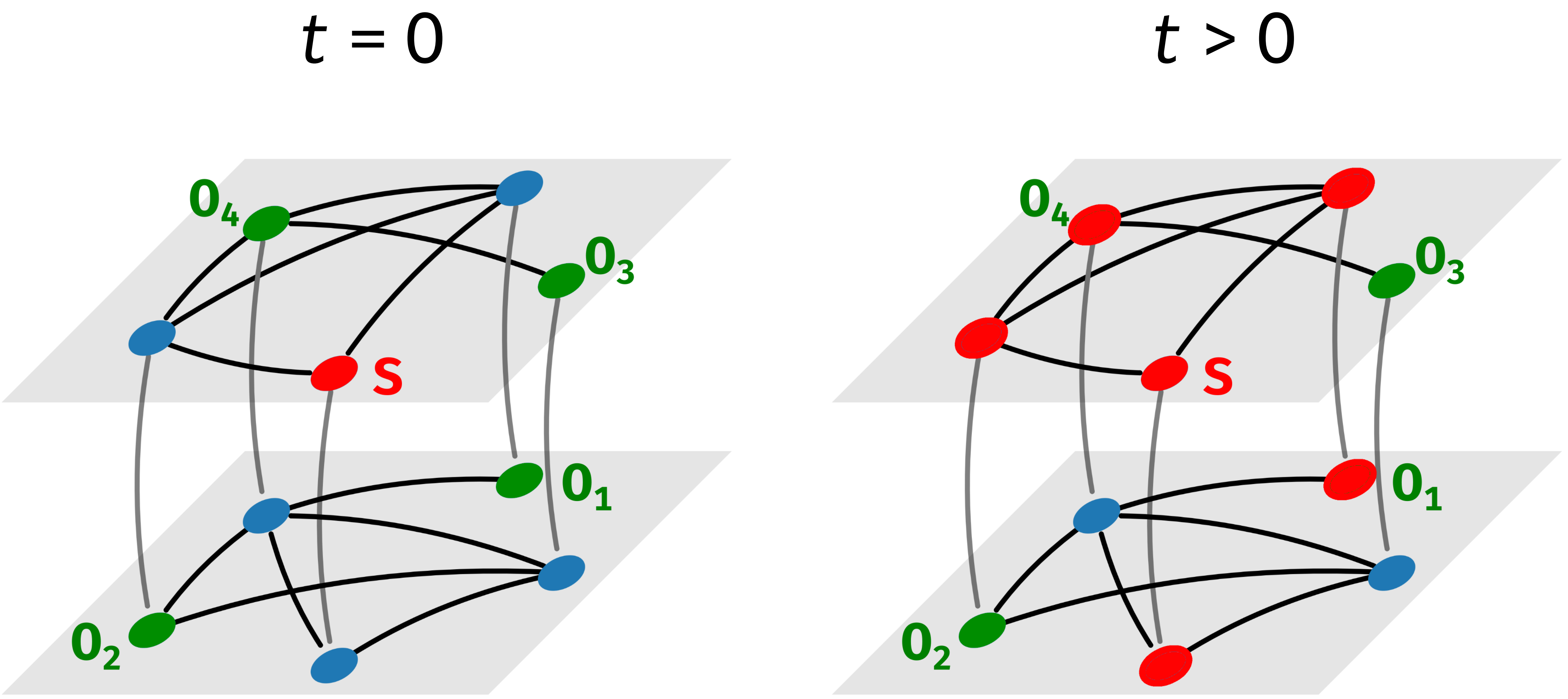


Goal

- Estimate the likelihood of being the source for each replica.
- Select the replica with the highest likelihood.

$$\hat{s} = \arg\max_{s \in V} \phi(s)$$

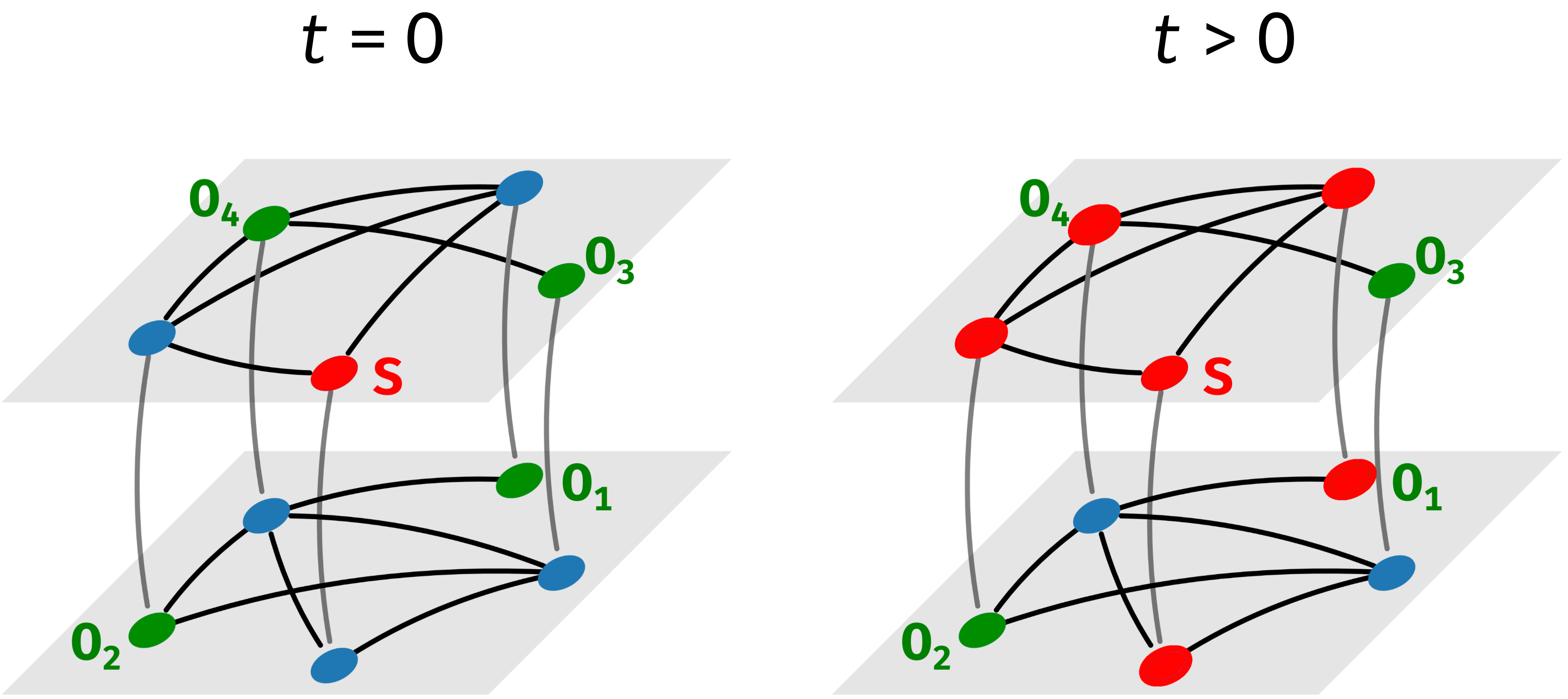
Observers-based maximum likelihood estimation



Assumptions:

- Topologies of all layers are known.

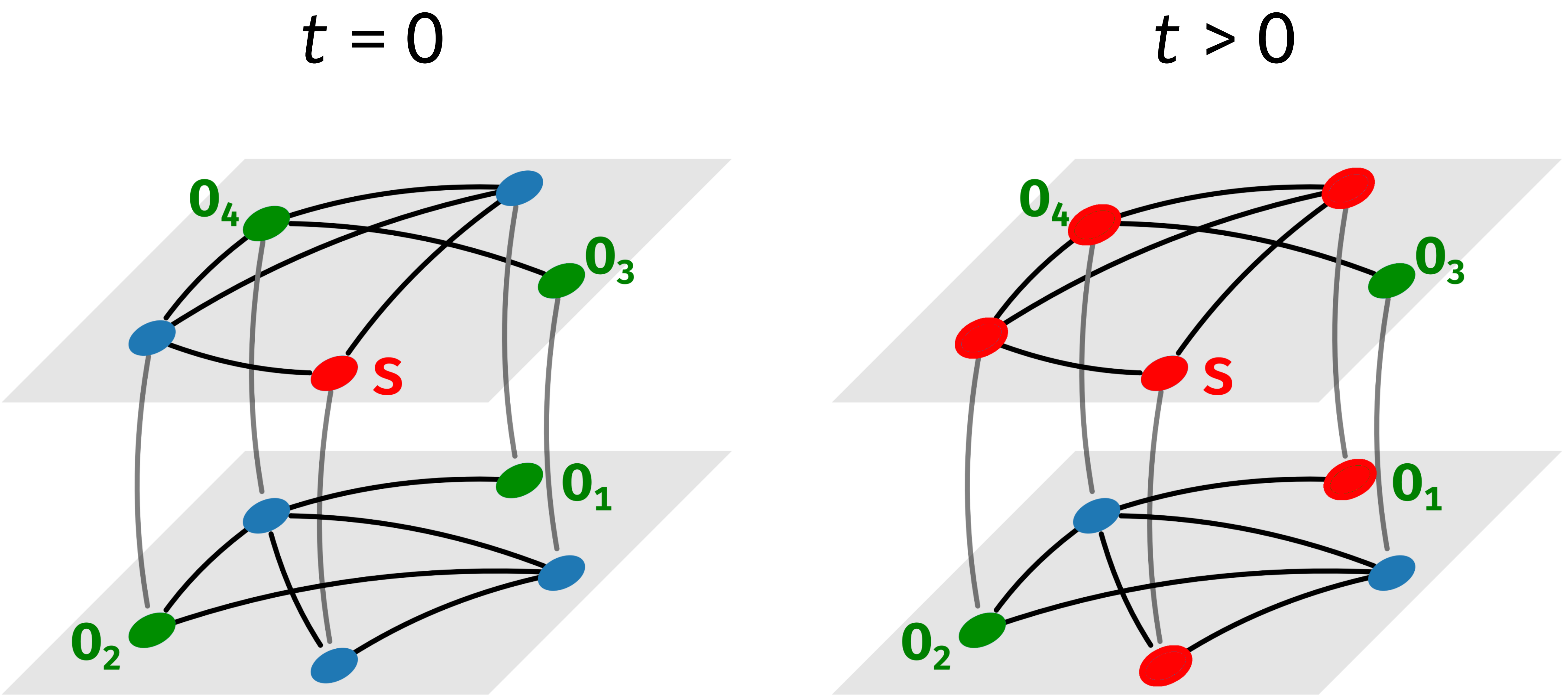
Observers-based maximum likelihood estimation



Assumptions:

- Topologies of all layers are known.
- Spreading parameters of all layers and interlinks are known.

Observers-based maximum likelihood estimation

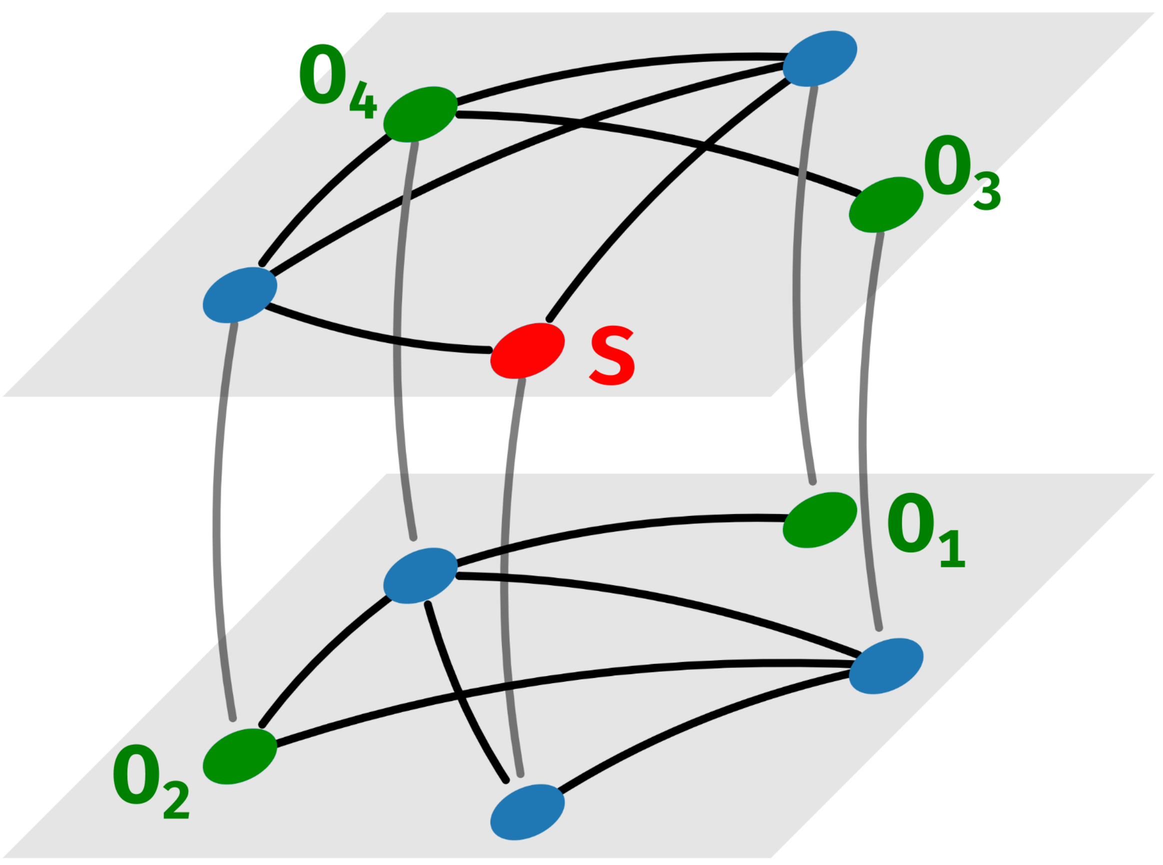


Assumptions:

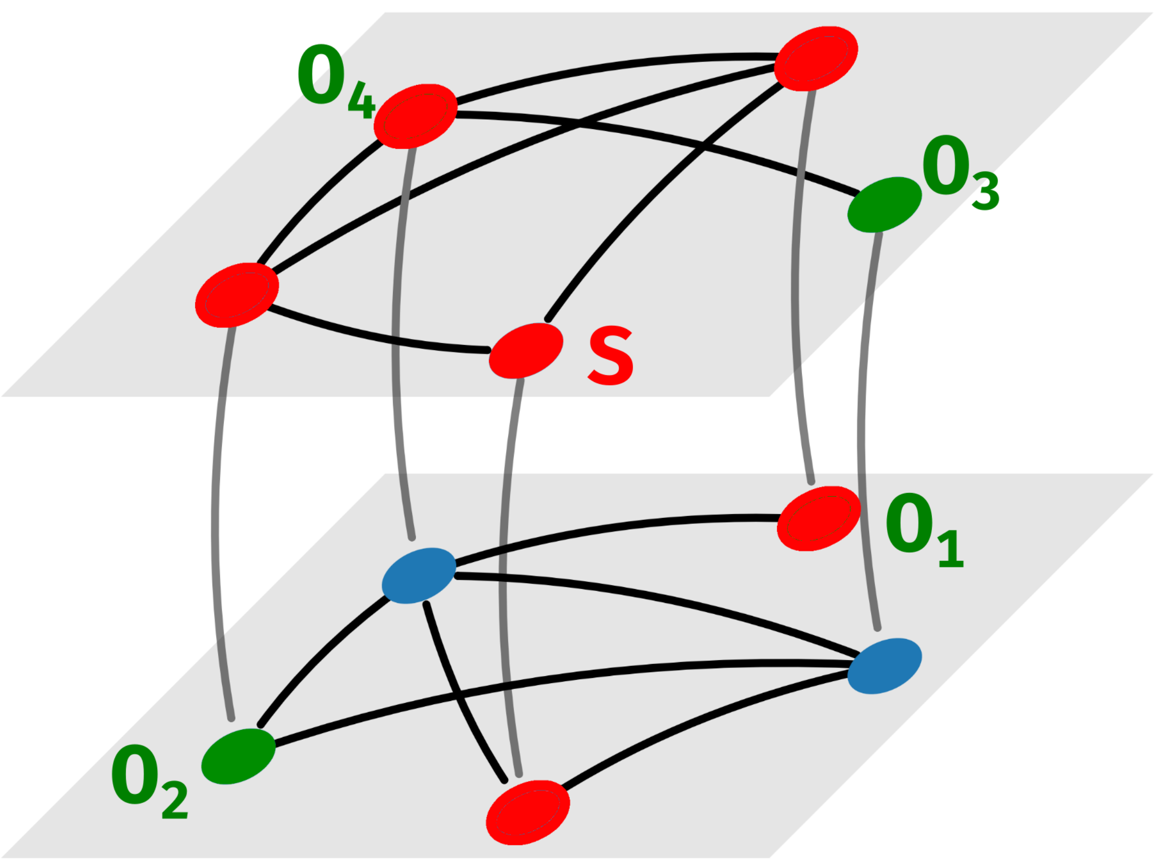
- Topologies of all layers are known.
- Spreading parameters of all layers and interlinks are known.
- Infection times of <u>some</u> replicas (called <u>observers</u>) are known.

Observers-based maximum likelihood estimation

$t = 0$



$t > 0$

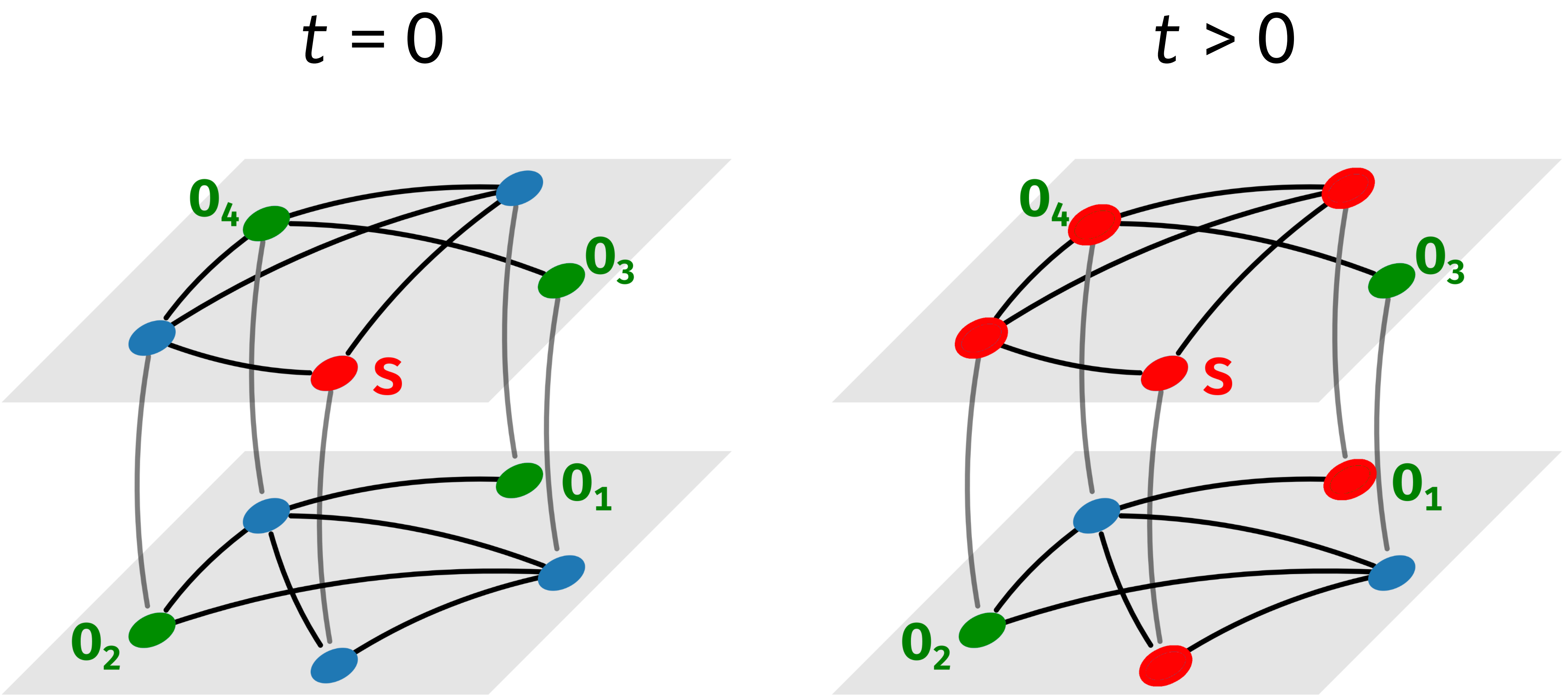


Assumptions:

- Topologies of all layers are known.
- Spreading parameters of all layers and interlinks are known.
- Infection times of <u>some</u> replicas (called <u>observers</u>) are known.
- Time delays on the links are i.i.d. random variables with Gaussian distribution.

Politechnika
Warszawska

Observers-based maximum likelihood estimation



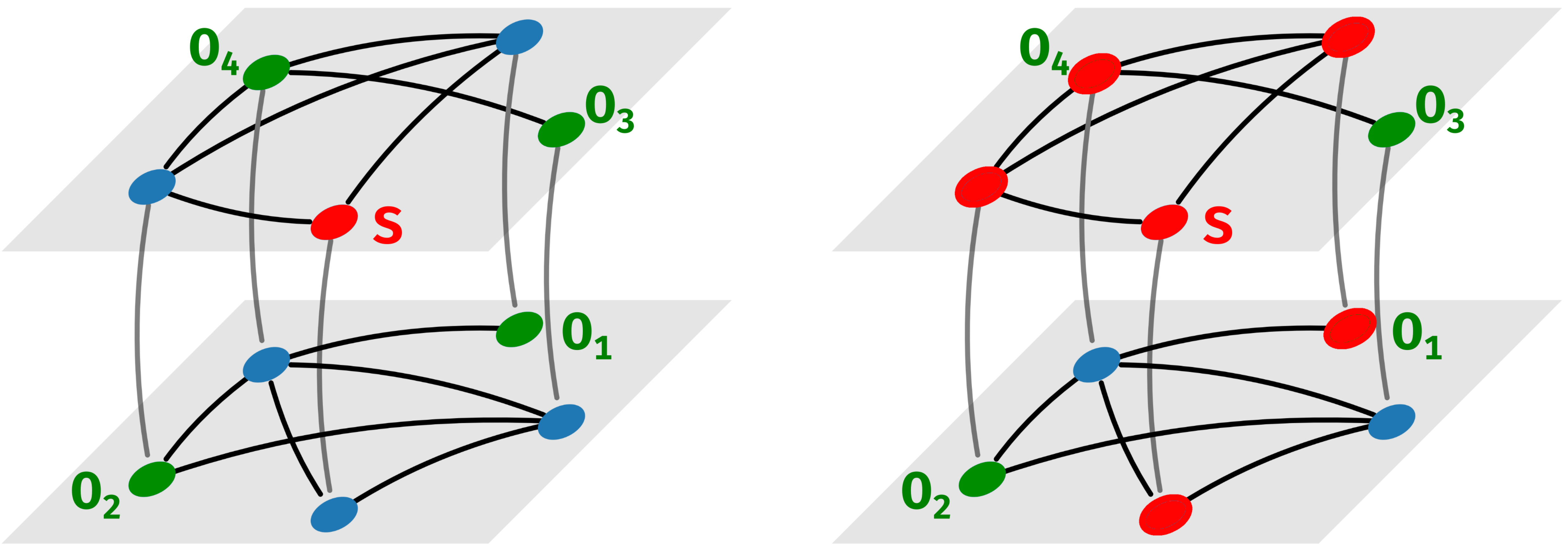
Assumptions:

- Topologies of all layers are known.
- Spreading parameters of all layers and interlinks are known.
- Infection times of <u>some</u> replicas (called <u>observers</u>) are known.
- Time delays on the links are i.i.d. random variables with Gaussian distribution.
- Information travels from the source to the rest of the nodes along the shortest paths.

Politechnika Warszawska

Observers-based maximum likelihood estimation

$t = 0$ $\qquad\qquad$ $t > 0$



$$\phi(s) = \frac{1}{\sqrt{(2\pi)^{K-1}|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{\mu}^{(s)})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{d}-\boldsymbol{\mu}^{(s)})}$$

Assumptions:

- Topologies of all layers are known.
- Spreading parameters of all layers and interlinks are known.
- Infection times of <u>some</u> replicas (called <u>observers</u>) are known.
- Time delays on the links are i.i.d. random variables with Gaussian distribution.
- Information travels from the source to the rest of the nodes along the shortest paths.

Politechnika
Warszawska

Experiment (scheme)

1. **Generate a multilayer graph (ER, BA or real structure).**

Experiment (scheme)

1. Generate a multilayer graph (ER, BA or real structure).
2. **Select random replica to be a source and run Susceptible-Infected model.**

Experiment (scheme)

1. Generate a multilayer graph (ER, BA or real structure).
2. Select random replica to be a source and run Susceptible-Infected model.
3. **Select random observers and calculate the likelihoods of being the source for all nodes.**

Experiment (scheme)

1. Generate a multilayer graph (ER, BA or real structure).
2. Select random replica to be a source and run Susceptible-Infected model.
3. Select random observers and calculate the likelihoods of being the source for all nodes.
4. **Find the rank of the true source in terms of the likelihood. The rank equal to 1 means successful identification (*prec* = 1). However, other low values of rank are also beneficial. The draws are possible and cause fractional precision (e.g. *prec* = 0.5, 0.3(3), 0.25).**

Experiment (scheme)

1. Generate a multilayer graph (ER, BA or real structure).

2. Select random replica to be a source and run Susceptible-Infected model.

3. Select random observers and calculate the likelihoods of being the source for all nodes.

4. Find the rank of the true source in terms of the likelihood. The rank equal to 1 means successful identification (*prec* = 1). However, other low values of rank are also beneficial. The draws are possible and cause fractional precision (e.g. *prec* = 0.5, 0.3(3), 0.25).

5. **Repeat steps 1-4 many times and compute the average precision and the Credible Set Size of order α (a quantile of order α of the rank).**

Experiments (types)

1. Two-layers problem (using ER and BA models).

Experiments (types)

1. Two-layers problem (using ER and BA models).
2. Many layers problem (using ER and BA models).

Experiments (types)

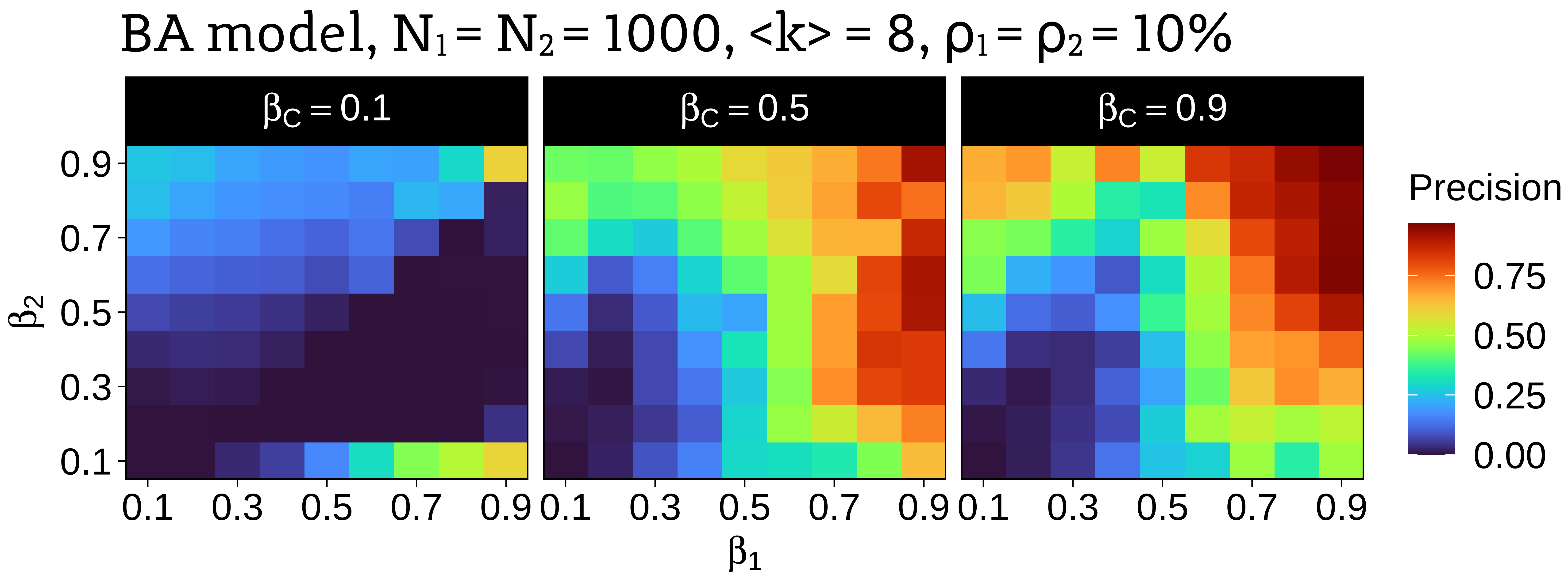
1. Two-layers problem (using ER and BA models).
2. Many layers problem (using ER and BA models).
3. Tests on the real multilayer structure (4 types of interactions between employees of the Department of Computer Science at Aarhus University)

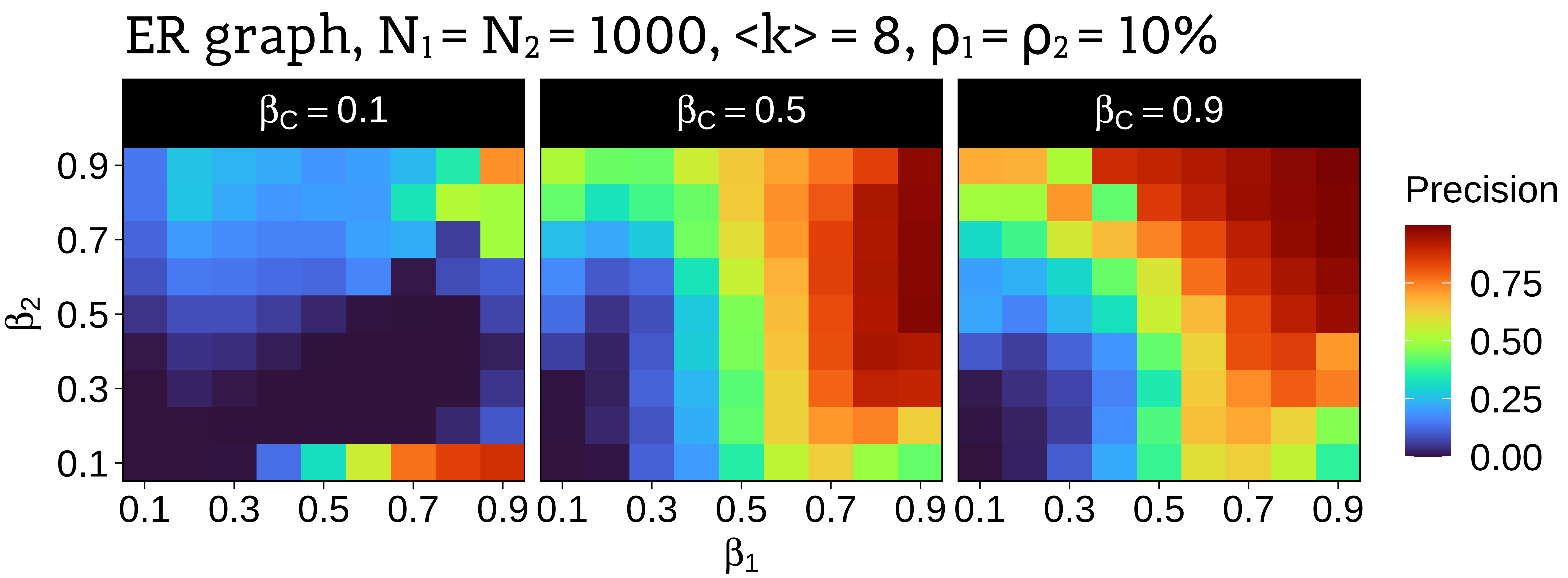
Layer	$\|V\|$	$\langle k \rangle$	k_{max}	n_c	$\langle l \rangle$	Diam.	C
Work	60	6.47	27	1	2.39	4	0.65
Lunch	60	6.43	15	1	3.19	7	0.70
Leisure	47	3.74	14	2	3.12	8	0.50
Facebook	32	7.75	15	1	1.96	4	0.54

Two-layers problem – different infection rates

The source is selected always from the replicas in **layer 1**.

A transition between **two modes** as a function of the crosslayer infection rate β_C.

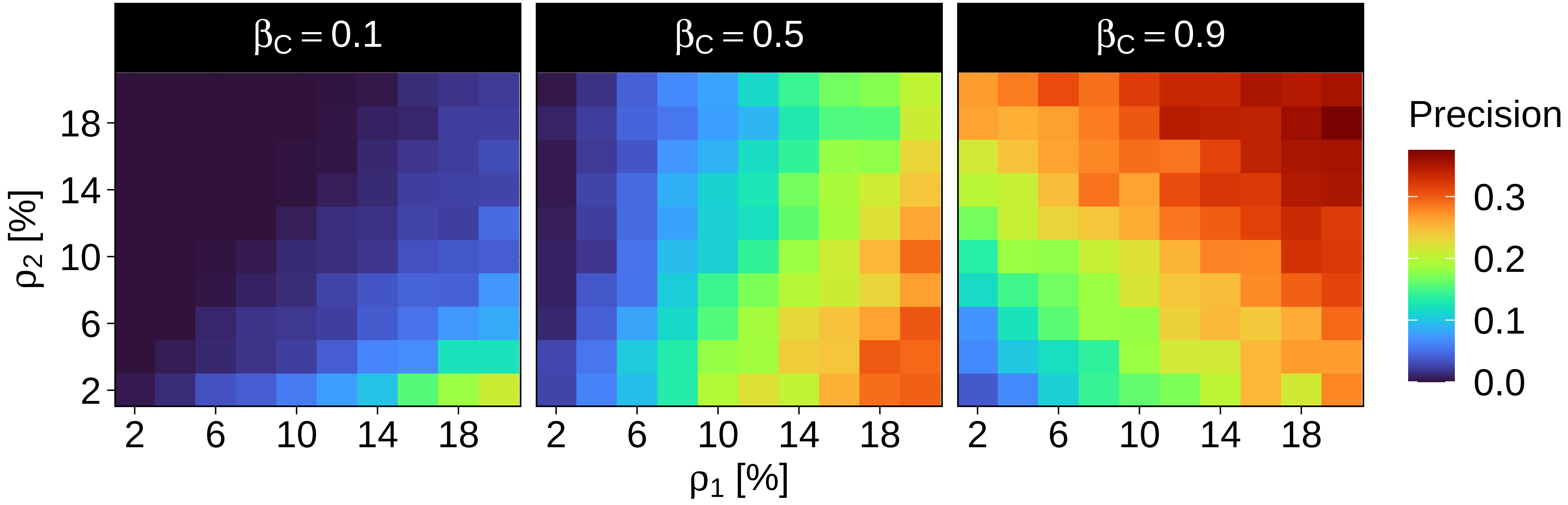




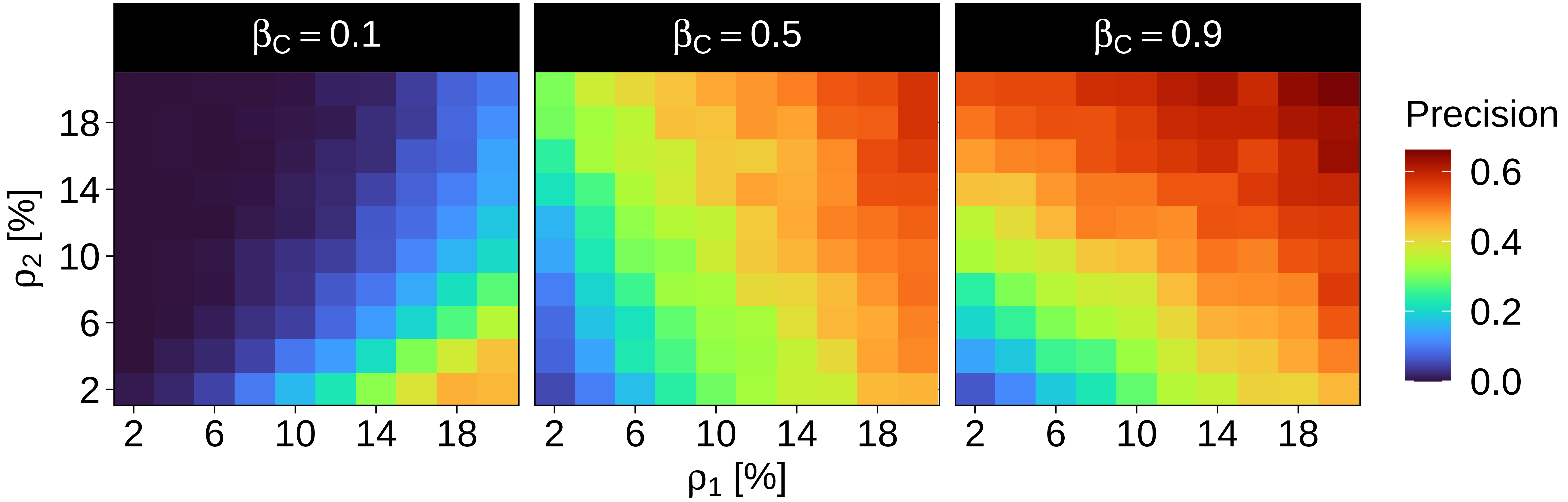
Politechnika Warszawska

Two-layers problem – different observers densities

BA model, $N_1 = N_2 = 1000$, <k> = 8, $\beta_1 = \beta_2 = 0.5$



ER graph, $N_1 = N_2 = 1000$, <k> = 8, $\beta_1 = \beta_2 = 0.5$



The source is selected always from the replicas in **layer 1**.

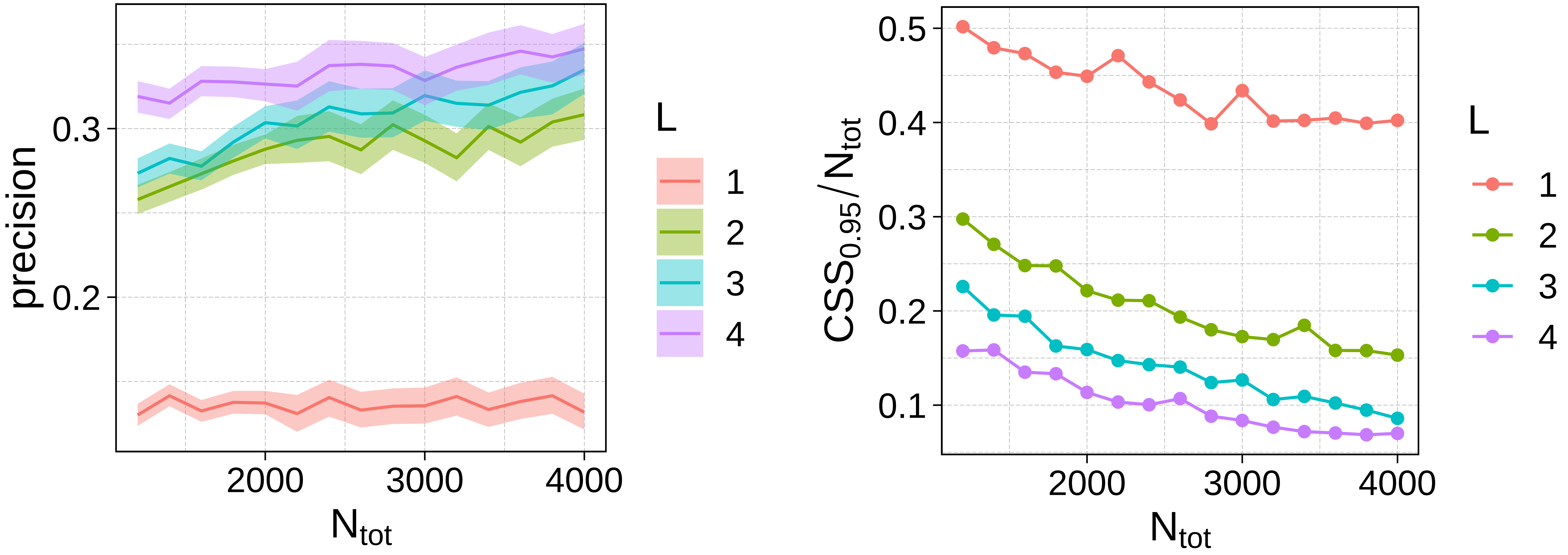
A transition between **two modes** as a function of the crosslayer infection rate β_C.

Politechnika Warszawska

Many layers problem – BA model, <k>=8, ρ=10%

$\beta_C = 0.8$

$\beta_i = 0.5$

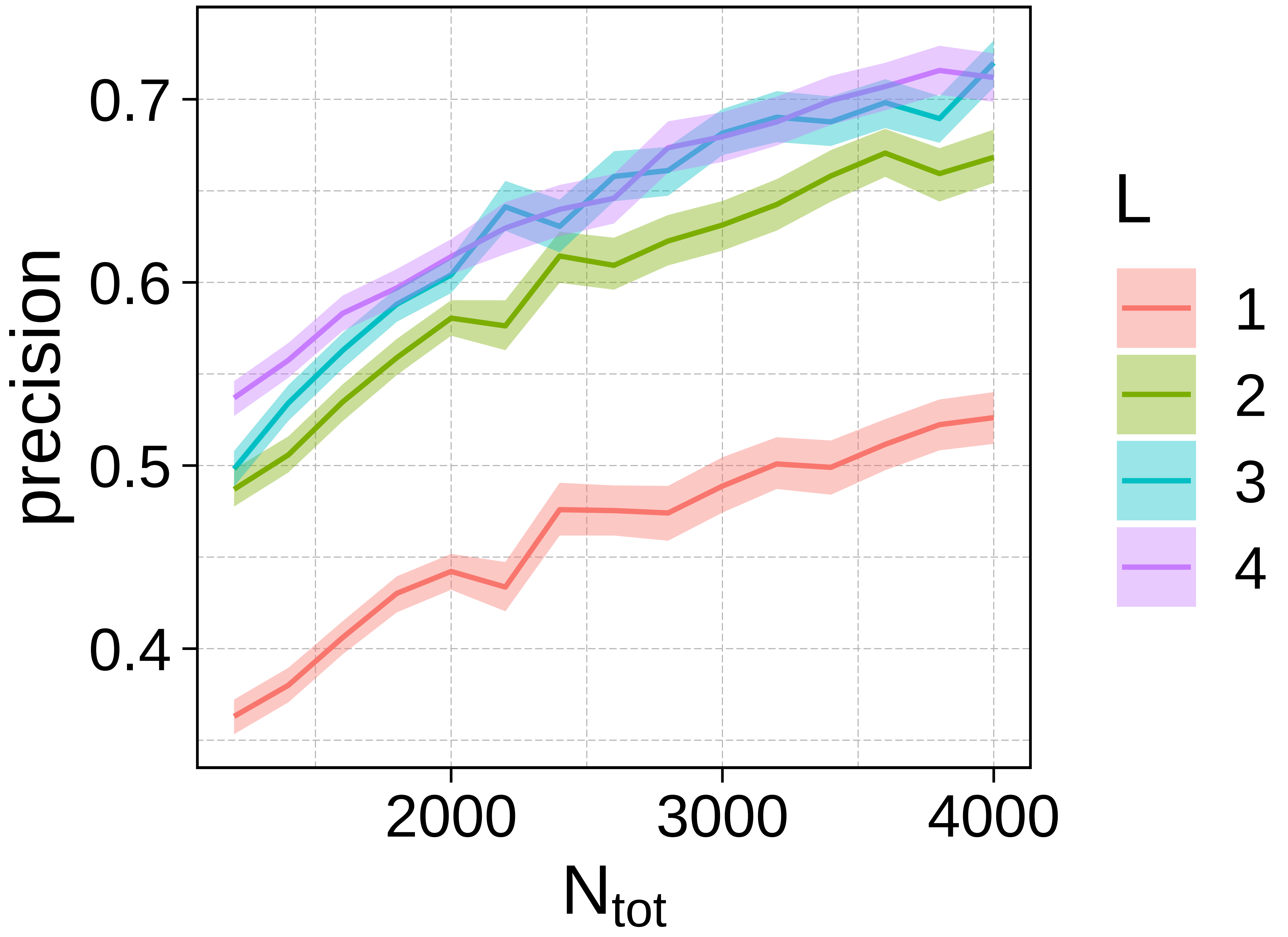


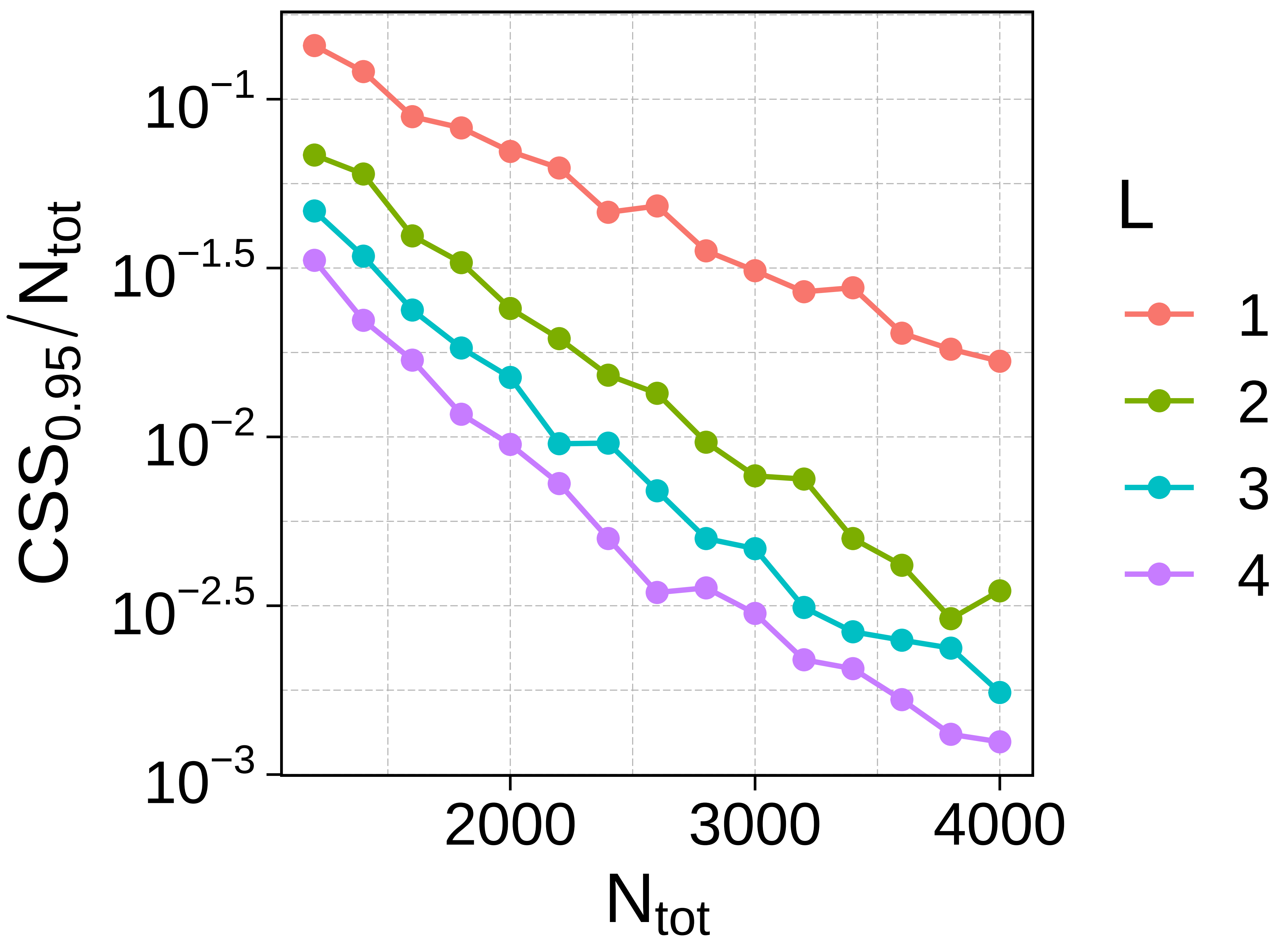
Politechnika Warszawska

Many layers problem – ER model, <k>=8, ρ=10%

$\beta_C = 0.8$

$\beta_i = 0.5$

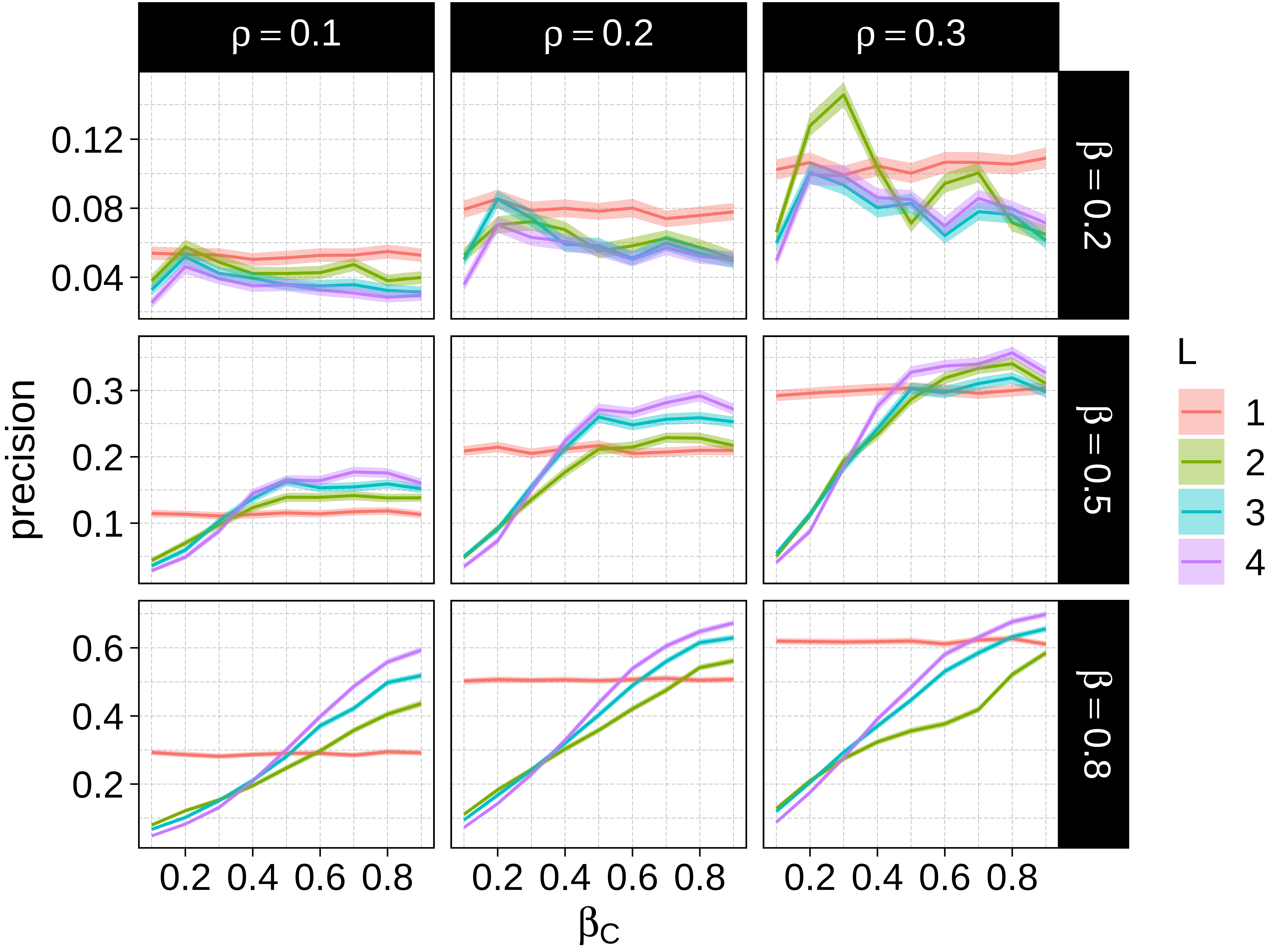




Politechnika
Warszawska

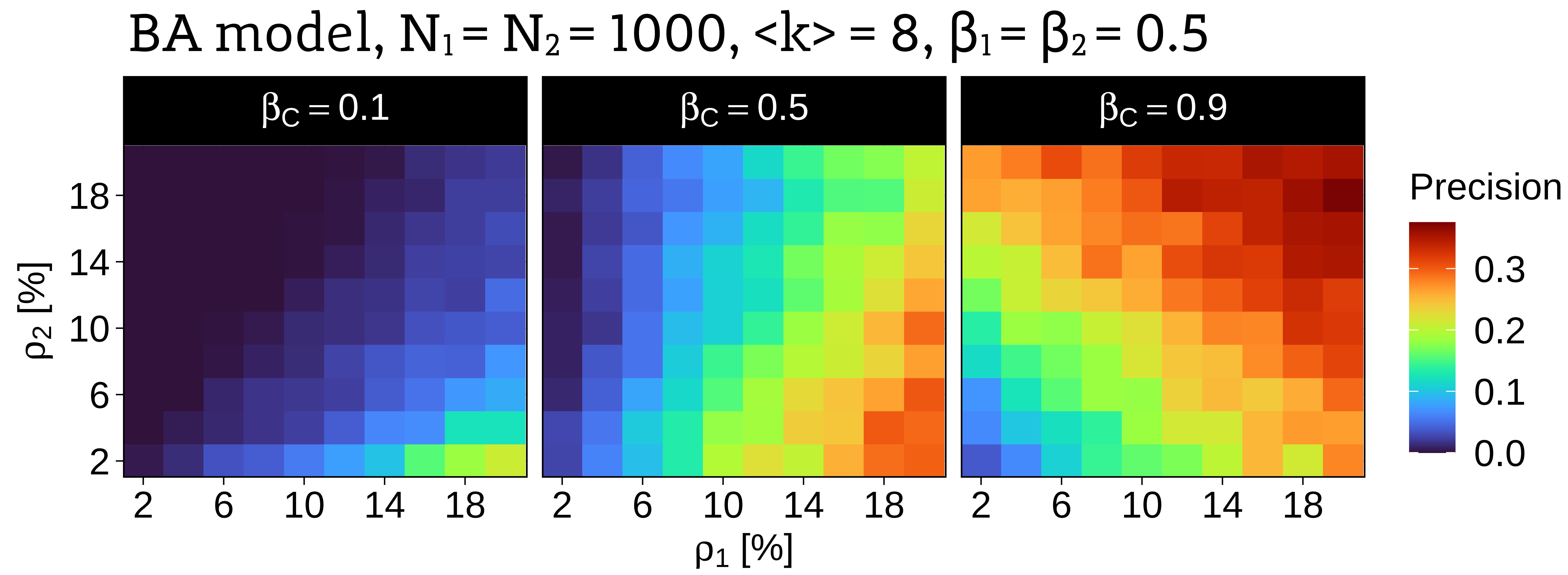
Aarhus University network

Layers are added from the largest (Work) to the smallest (Facebook).

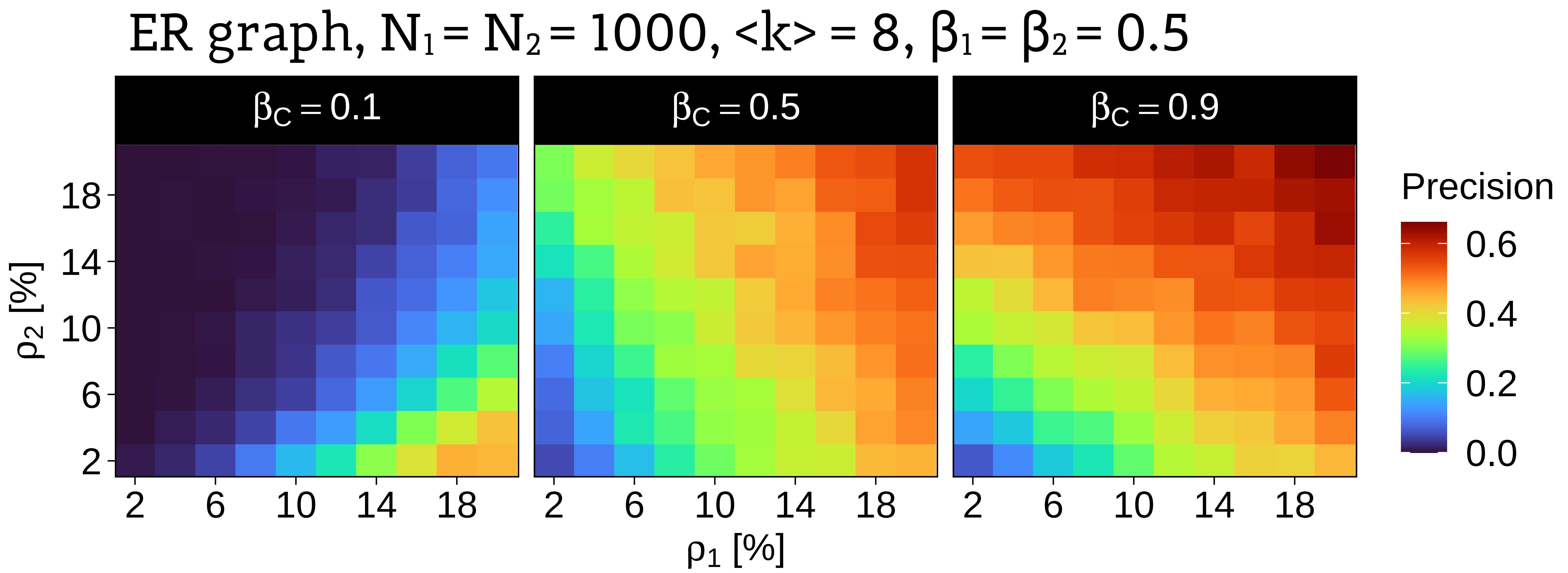


Politechnika Warszawska

Two-layers problem – different observers densities



The source is selected always from the replicas in **layer 1**.



A transition between **two modes** as a function of the crosslayer infection rate β_C.

Politechnika Warszawska

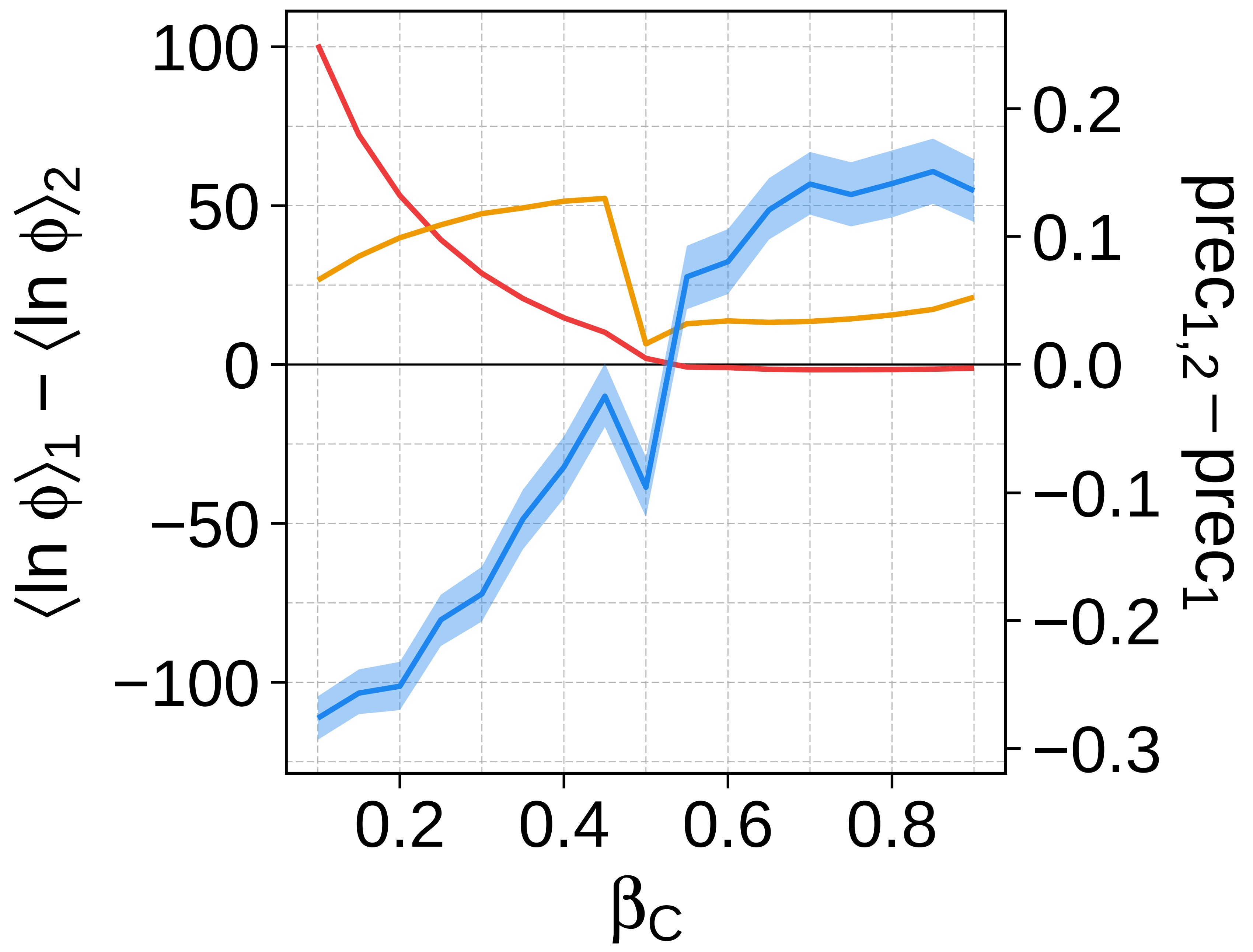
A transition between the interference mode and synergy modes (for two layers)

A difference between average log-scores in the first and the second layer using the observers:

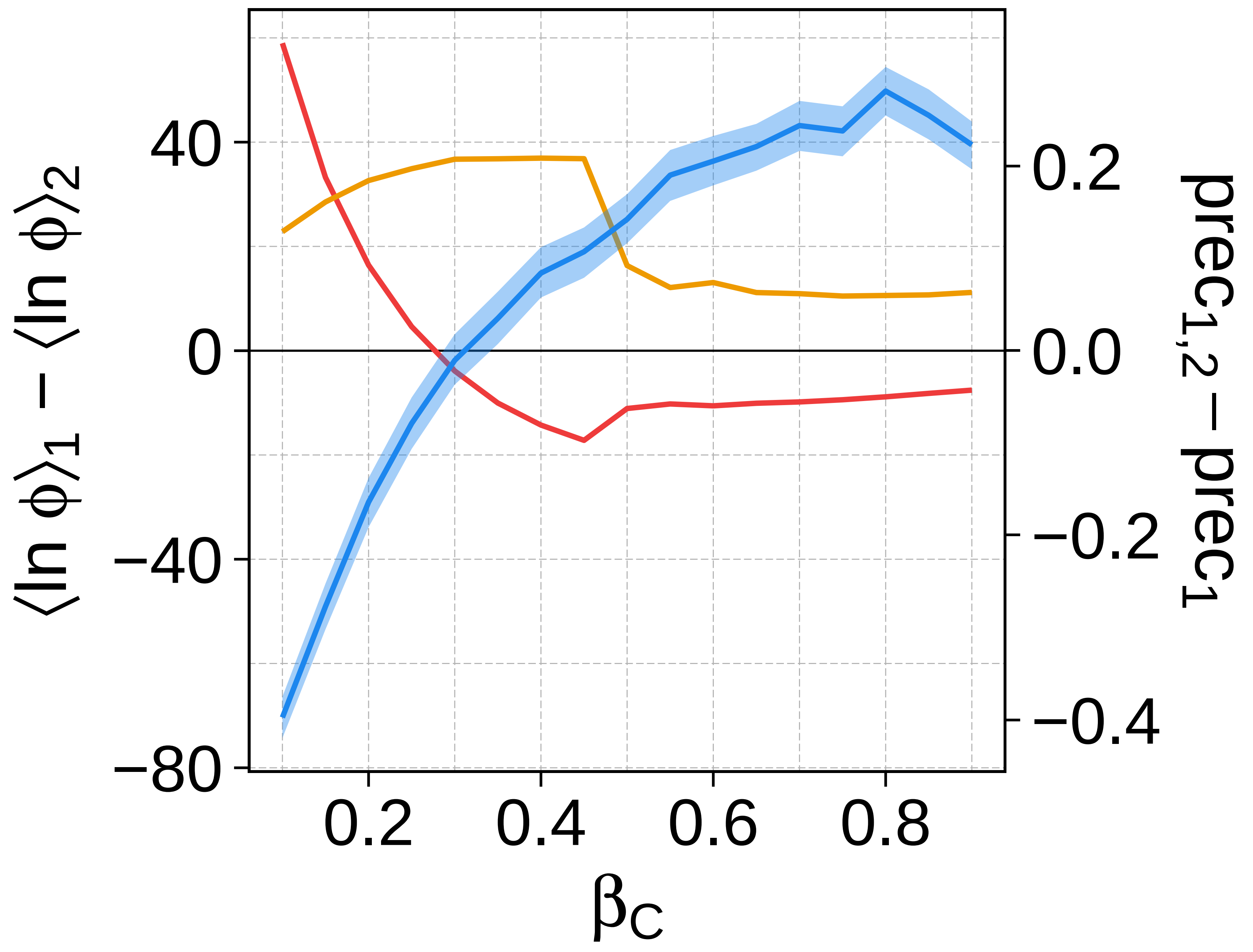
- **from both layers,**
- **only from the first layer, where the source is.**

In blue: a difference between the average precision of source localizaton using the observers from both layers and only the first layer.

BA, $N_1 = N_2 = 500$, <k> = 8, $\beta_1 = \beta_2 = 0.5$



ER, $N_1 = N_2 = 500$, <k> = 8, $\beta_1 = \beta_2 = 0.5$



Politechnika Warszawska

Summary

- The multilayer topology makes it easier to locate the source if the layers are strongly coupled, and we have detectors in both layers.
- In the case of weak coupling, observations from the sourceless layer may degrade detection performance.
- The mean score is higher in the layer where the source is, which allows for easy identification of that layer.

Literature

1. Paluch, R., Lu, X., Suchecki, K., Szymański, B. K., & Hołyst, J. A. (2018). *Fast and accurate detection of spread source in large complex networks*. Scientific Reports, 8(1), 2508. https://doi.org/10.1038/s41598-018-20546-3

2. Paluch, R., Gajewski, Ł. G., Hołyst, J. A., & Szymanski, B. K. (2020). *Optimizing sensors placement in complex networks for localization of hidden signal source: A review*. Future Generation Computer Systems, 112, 1070–1092. https://doi.org/10.1016/j.future.2020.06.023

3. Paluch, R., Gajewski, Ł. G., Szymański, B. K., & Hołyst, J. A. (2021). *Impact of interactions between layers on source localization in multilayer networks*. Physica A (in print).