Physica A 582 (2021) 126238

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Impact of interactions between layers on source localization in multilayer networks

Robert Paluch [a], Łukasz G. Gajewski [a], Krzysztof Suchecki [a], Janusz A. Hołyst [a,b,*]

[a] Faculty of Physics, Warsaw University of Technology, Koszykowa 75, Warsaw, 00-662, Poland
[b] ITMO University, Kronverkskiy Prospekt 49, St. Petersburg, 197101, Russia

ARTICLE INFO

Article history:
Available online 6 July 2021

Keywords:
Source localization
Multilayer networks
Susceptible–infected model
Maximum likelihood estimation

ABSTRACT

Nowadays it is not uncommon to have to deal with dissemination on multilayer networks and often finding the source of said propagation can be a crucial task. We examine the issue of locating the source of Susceptible–Infected spreading process in a multilayer network using the Bayesian inference and the maximum likelihood method established for general networks and adapted here to cover multilayer topology. We show how its accuracy depends on network and spreading parameters and find the existence of two parameter ranges with different behavior. If inter-network spreading rate is low, observations in different layers interfere, lowering accuracy below that of relying on single layer observers only. If it is high, on the other hand, observations synergize, raising accuracy above the level of single layer network of the same size and observer density. We also show a heuristic method to determine the case in a system and potentially improve accuracy by rejecting interfering observations. This paper is dedicated to the memory of Professor Dietrich Stauffer, who was a pioneer in new approaches in statistical physics and its interdisciplinary applications.

Dedication and memoir

We dedicate this paper to Professor Dietrich Stauffer (1943–2019), who was one of the pioneers in the new studies of statistical physics, computational physics as well as econophysics and sociophysics. His work impacted a vast number of researches all over the world. Below is a short memoir from one of us (JAH), who had the pleasure to work with him directly.

Dietrich Stauffer was a great scientist as well as a fantastic Colleague and I am very proud I was able to collaborate with him on many occasions. Our first interaction was at the Winter School in Lądek Zdrój (Poland) in February 2000. Dietrich impressed me with his lecture and tutorials on programming Monte Carlo simulations for large Ising models of hypercubic lattices in d-dimensions. He presented a Fortran code that looked rather simple but was very fast in using the computer's memory in a highly effective manner. Subsequently, I learnt that for several years Dietrich had been improving his codes to perform simulations of large-scale Ising systems [1].

At the same venue I presented my model of strong leader for opinion dynamics [2,3] that was based on social impact theory and Dietrich invited me to submit a review paper about this model for the *Annual Reviews of Computational Physics*. As the *Volume Editor* Dietrich was very strict and fast-paced, he did not accept any amendments to deadlines and made the editorial decisions very swiftly. I wish I could act in this way as the present *Main Editor* in *Physica A*! Most likely because

* Corresponding author at: Faculty of Physics, Warsaw University of Technology, Koszykowa 75, Warsaw, 00-662, Poland.
E-mail address: jholyst.physica@gmail.com (J.A. Hołyst).

https://doi.org/10.1016/j.physa.2021.126238

of his sharp and close supervision we managed to finish this review with Krzysztof Kacperski and Frank Schweitzer in only few months [4].

In Fall 2001, I was working with Dirk Helbing in Dresden and used this opportunity to visit Dietrich in Cologne for a weekend. In fact, my Ph.D. student (Agata Aleksiejuk, now Prof. Agata Fronczak) was being hosted by Dietrich as a DAAD fellow as such I was expecting that we would shortly summarize her progress and then enjoy the nice weather and sightseeing of Cologne's architecture. Turned out my assumption was completely wrong.

Dietrich picked me up at the airport and while we were in the shuttle he inquired what I would like to do during these two days - *Janusz, you came here to work and so we need to fix a research plan*. At the time complex networks were a new concept thus I suggested we could study Ising model in a scale-free network. *How do you implement such dynamics and why could this be an interesting problem?* Dietrich asked. I replied that a complex network could be treated similarly as periodic lattices with nearest neighboring interactions, where the proximity is defined by the presence of direct links between nodes. At that time it was also an open question, if any magnetic structures could be observed in such a non-periodic system.

Before our bus arrived at the center of Cologne, we had a research plan. Dietrich took me to his Institute (Agata was working there in the library) and the three of us began brainstorming, how to implement so-called Kertesz algorithm of Barabási–Albert network for Ising simulations. Before the end of the day first results of simulations came in: there was a spontaneous ferromagnetic order in low temperatures. We decided to celebrate this finding with a dinner in a restaurant.

On Sunday morning, we returned to the Institute to find out that the overnight simulations run at Dierich's computer indicate that the critical temperature for this system varied depending on the networks size. It was a very unexpected result and we immediately started to search for errors in the computer's code. Some minor bugs were found but their removal did not substantially affect the outcome. Around lunch time Dietrich noticed that the critical temperature was increasing as a logarithm of number of spins. Finally, I had to leave Cologne but in a few weeks the preprint was ready. After the paper had been published [5] I discovered that similar studies were ongoing in other groups but our team was the first group who had observed these effects.

Over the years, I was meeting Dietrich frequently at the econophysics and sociophysics conferences, where we discussed among other topics the scaling conditions for critical exponents in random directed percolation that I used to model for collective bank bankruptcies [6,7]. In 2008, Dietrich came to Warsaw for a few weeks to work with me on models of social dynamics [8,9].

Dietrich was well known for his witty and ironic sense of humor that transpired not only in research but also in his comments on politics, history and football. Below are two of his e-mails to me.

> Dear Janusz,
>
> [9 Sept. 2013]
>
> *our newspaper reported that some German sold PhD certificates of Warszawa Technical University for up to 15,000 Euro per doctorate. There is no hint in the report that this university was involved. But if it was and shared the money with him, is this the reason that Poland got much better though the 2009 economic crisis than the rest of the European Union. Dietrich*
>
> [14 July 2014]
>
> *Good news: Footballer Miroslaw Kloze (= Klose) from Opole (Slask), exiled to Lazio Roma (Italy), during a holiday in Brazil with his German friends Podolski, Ozil, Khedira, Boateng et al. made a new world record with 16 goals in his four world championships. Now he is 36 years old and presumably will not participate again four years from now.*
>
> *Bad news: Erika Steinbach wants to retire. Thus there is little hope left for you that she will be crowned queen of Poland(-Lituania?) on the Wawel. What can Poland do now ? Ask Tsar Putin to accept you as "prewisliski kraj" in Rodina-mat' Rossia, like the Crimea ? Dietrich * * * **

No doubt, Dietrich's curiosity contributed tremendously to many branches of physics and interdisciplinary research and inspired many scientists. Nevertheless, in my memory he remains also as an open-minded individual with a deep interest in understanding social interactions and other cultures.

1. Introduction

Sharing and spreading information is one of the cornerstones of civilization. Not all information is desirable however, and some, such as misinformation or conspiracy theories can have a detrimental effect on the society as a whole. Therefore, it is of utmost importance to research all aspects of such processes so that we can develop tools to address potential risks appropriately. In practice, one of the fundamental questions is that of the true origin of a spread which often is very important as it allows for developing or implementing appropriate preventive measures and deepen the understanding of the spread itself. This is not very dissimilar to the well-known problem of locating a "patient zero" in the field of epidemiology. The question of how to find the source of information or a rumor is not new and we have seen a large

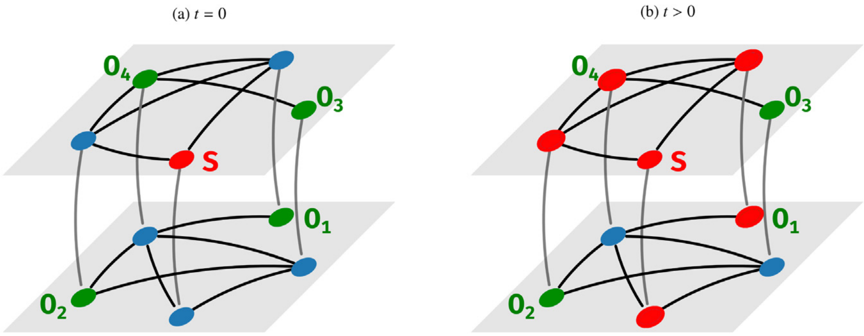

Fig. 1. Schematic illustration of studied systems. Two snapshot of a propagation process are presented one in the very beginning $t = 0$ (left) and the other at some later time $t > 0$ (right). The red nodes are infected with the one labeled S being the source. Blue and green nodes are susceptible where green ones (and labeled with O_i) are the observers. This example shows the idea of the setting used in our experiments - a multilayer structure yet the states are not shared between images of nodes. E.g. both O_1 and its image are infected in the right panel, however, while O_4 has its replica has not received the signal yet. Additionally it is worth noting that an observer in one layer is not necessarily an observer in the other.

amount of research devoted to this topic. Some methods are based on a single snapshot of nodal states at a given moment in time. If the snapshot covers all nodes, it is a complete observation [10–13], but sometimes only a subset of nodes is required [14–18]. Another type is detector-based algorithms, where a small subset of nodes, called observers or detectors, is monitored all the time [19–26]. From these observers, we know the exact time when they received a message. More recent works aim at relaxing the assumptions about the process and its parameters [27,28] to allow finding a source of a signal without prior knowledge of the actual spreading process. There have been also works considering source finding problem on a variable, temporal topology of connections [29,30]. Despite all the effort put into the research of source localization, to our knowledge, this issue has not yet been studied in the context of multilayer networks. In this work we address this problem, since it has been shown that such structures happen to be quite prevalent in our society [31–40] and the nature of spreading processes on them has been studied extensively [41–48].

The organization of the article is as follows. Section 2 presents the definitions of multilayer graph (Section 2.1) and propagation process (Section 2.2). This section also contains the descriptions of the source localization method (Section 2.3) and evaluation metrics (Section 2.4). The results of numerical simulations on synthetic and real networks are shown in Section 3. We use the multilayer versions of Erdős–Rényi (ER) and Barabási–Albert (BA) models and investigate them in cases of two (Section 3.1) and more than two layers (Section 3.2). As an example of real multilayer structure we use the data collected by the Department of Computer Science at Aarhus University among the employees [49] (Section 3.3). Section 4 discusses and summarizes the obtained results.

2. Preliminaries

2.1. Multilayer graphs

In this paper we consider a multilayer graph with L denoting number of layers. Each layer has n_l nodes and m_i edges where $i \in [1, 2 \ldots L]$. The total number of nodes in the network is $n_{tot} = Ln_l$. Each layer has its own topology, but they can be potentially correlated (more on that later). The users are present in different layers as *replicas*, that have separate internal states and are connected to all other replicas of the same user. It is conditioned by the fact that people can exhibit distinct preferences, opinions, and behaviors in different social networks. It follows that sometimes they may not relay information learned in one network to another or do so with a delay. See Fig. 1 for an example.

We conduct our studies on two well-known synthetic network models: Erdős–Rényi (ER) and Barabási–Albert (BA). As mentioned before each layer is independent, i.e. we construct L realizations of a given graph model and couple them accordingly with interlayer links. While every layer is different from others the degree distributions in the BA model are highly correlated — a hub in one layer is most likely a hub in another. Moreover, we verify our findings from the network models mentioned above on the real-world multilayer network depicting four types of contact among the employees of Aaarhus University.

2.2. Spreading on multilayer structure

We use an agent-based version of Susceptible–Infected model [50] to simulate the spread across the graph. According to this model, an agent may be in one of two states, susceptible (S) or infected (I). At the beginning of the simulation, only one agent is infected — it is the true source. In the next steps the infected agents interact with their neighbors which as a result may change the susceptible nodes into infected with probability β per time step (which is called an infection

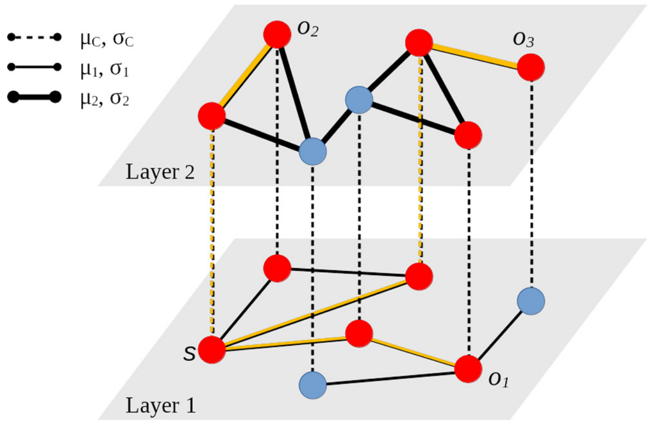

Fig. 2. Schematic illustration of the information traversal dynamics. The node labeled S is the source of the spread and the nodes o_1, o_2, o_3 are observers. Red nodes already received the information. Each layer has its own set of propagation properties (mean - μ - and standard deviation - σ - of the traversal time) and the interlayer coupling is also independent with its respective parameters. The orange color on certain edges indicates the shortest paths from the source to the observers.

rate). To simplify the tracking of propagation paths and infection times, the dynamics of our model is synchronous, which means that at every time step, all infected nodes try to pass the infection simultaneously to their all susceptible neighbors.

In the case of a multilayer structure, we assume that the infection rate β depends on the layer. Additionally, we allow replicas of an agent to have different states in each layer, which corresponds to a situation when a social media user shares an information on one platform but does not on the others. We define interlayer infection rate as the probability in each step that an agent will transfer information from one layer he is present in to another, which corresponds to infection of a replica in one layer by the replica in another. The infection rate can be therefore written as a matrix $\boldsymbol{\beta} = [\beta_{kl}]$ where k is the layer of the infecting node and l is the layer of the node being infected. The diagonal values correspond to the intralayer infection rates, while non diagonal correspond to the interlayer infection rates. We assume that the interlayer infection rates are all the same $\beta_{k \neq l} = \beta_C$, while we can denote intralayer rates by a single index $\beta_{kk} \equiv \beta_k$.

We can write the equation for the probability of a susceptible agent i in a layer l becoming infected in next time step as

$$P_i^{(l)}(t+1) = 1 - \prod_{j \neq i} A_{ij}^{(l)} \left(1 - s_j^{(l)}(t)\beta_l\right) \prod_{k \neq l} \left(1 - s_i^{(k)}(t)\beta_C\right) \tag{1}$$

where $s_j^{(l)}(t)$ is the current state of an agent j in the layer l, with value 0 for Susceptible and 1 for Infected, and $A_{ij}^{(l)}$ is an adjacency matrix in the layer l. The first product is due to the intralayer propagation, while the second product is due to the interlayer propagation between replicas of a node i in different layers.

2.3. Source location

We estimate the source location using detectors-based maximum likelihood estimator. This method was first introduced by Pinto et al. [19] for single layer graphs. In this work we extend this algorithm to be able to locate the source in the multilayer structures with different diffusion properties for each layer. We assume that the distributions of delays on the links in each layer, which describe a spreading process, have finite means $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_L, \mu_C]$ and variances $\boldsymbol{\sigma}^2 = [\sigma_1^2, \sigma_2^2, \ldots, \sigma_L^2, \sigma_C^2]$, where μ_C and σ_C^2 describe the distribution of delays on all interlinks between layers. Moreover, the method requires that all these means and variances are known, as well as all links in each layer. What we do not know is the full history of the process. We can only monitor the states of some preselected replicas $o_i \in O$, called observers. Please note, that observing a replica in some layer does not mean that we monitor the states of corresponding replicas in the others layers. In that sense, we consider replicas to be for all functions and purposes separate nodes, so one observer assignment means one replica in one layer. We treat the network as if it was a single layer, but with different delay distributions on links defined by different layers. Being able to monitor observers, we know exactly when a specific message (infection) arrived at the observer node. From infection times reported by observers we construct an observed delay vector $\mathbf{d}$:

$$\mathbf{d} = (t_2 - t_1, t_3 - t_1, \ldots, t_b - t_1)^T \tag{2}$$

where b is the number of observers (available budget for the observation process), t_i is an infection time of observer $o_i \in O$, t_1 is the infection time of a *reference observer* o_1 and T stands for transposition (we will express the equations in terms of matrix algebra, hence column and row vectors are distinct). Since we do not know when the initial message appeared, we have to make all times relative to some reference point, and we choose it as the time at which the information arrives at the reference observer.

Then, for each node $v \in V$ in the whole network we construct a tree $\mathcal{T}_v$ from the shortest weighted paths (which may contain inter- and intralinks) between v and all observers $o_i \in O$. The weights of the links depend on the layer and are given by the vector $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_l, \mu_C]$, see Fig. 2 for illustration. The rest of the following computations for node v is performed on tree $\mathcal{T}_v$, not on the whole graph $\mathcal{G}$. This is an approximation, relying on that most non-lattice networks being locally treelike. Existence of loops make exact analytical predictions of time distributions much harder [25].

The method hinges on the comparison between the distribution of expected arrival times and the actual measurement. The Bayesian probability $P(s = v|\mathbf{d})$ for a specific node v being the true source s given observed delay vector is proportional to the probability of obtaining observed times $\mathbf{d}$ given the node v is the source $P(\mathbf{d}|s = v)$ (from the expected delay distribution), under the assumption that all nodes are equally likely to be the source a priori ($P(v = s) = 1/N$).

$$P(v = s|\mathbf{d}) = P(\mathbf{d}|v = s)\frac{P(\mathbf{d})}{P(v = s)} \tag{3}$$

where $P(\mathbf{d})$ is constant, because we only need the estimate for the actual $\mathbf{d}$, so the factor is the same for all potential sources s, and $P(v = s) = 1/N$.

To obtain the likelihood for node v of being the source we need to estimate the distribution $P(\mathbf{d}|v = s)$ by computing the expected delay vector $\boldsymbol{\mu}_v$ and the covariance matrix $\boldsymbol{\Lambda}_v$:

$$\boldsymbol{\mu}_v[i] = |\mathcal{P}(v, o_{i+1})|_\mu - |\mathcal{P}(v, o_1)|_\mu \qquad i = 1, 2, \ldots b - 1, \tag{4}$$

$$\boldsymbol{\Lambda}_v[i, j] = |\mathcal{P}(o_{i+1}, o_1) \cap \mathcal{P}(o_{j+1}, o_1)|_{\sigma^2} \qquad i, j = 1, 2, \ldots b - 1, \tag{5}$$

where $\mathcal{P}(v, o_i)$ denotes the path (a set of links) between nodes v and o_i in the tree $\mathcal{T}_v$, while $A \cap B$ means a set of shared links between paths A and B. The operators $|\,\mathcal{P}\,|_\mu$ and $|\,\mathcal{P}\,|_{\sigma^2}$ denote respectively summing up the mean delays or the variances of delays on links in the path $\mathcal{P}$, not simply length of the path, since links in different layers have different delays.

Finally we compute a *score* S_v for each node v and use maximum likelihood rule to determine the most probable source of the epidemic $\hat{s}$:

$$S_v = \boldsymbol{\mu}_v^T \boldsymbol{\Lambda}_v^{-1}(\mathbf{d} - 0.5\boldsymbol{\mu}_v) \tag{6}$$

$$\hat{s} = \arg\max_{v \in V} S_v \tag{7}$$

The score is maximally simplified strictly increasing function of actual probability $P(\mathbf{d}|v = s)$ that takes form of a multivariate normal distribution, which is another approximation making analytical calculations feasible [19].

While the method is formulated for normally distributed delays, many of the spreading phenomena will follow more closely the characteristics of Susceptible–Infected that we decided to use in our investigation. This means that delays are actually from geometric distribution, in conflict with assumptions of the method. We accept this discrepancy and following decrease in accuracy, given the unsolved problems such distribution will pose to estimate the form and parameters of $P(\mathbf{d}|v = s)$. The mean and variance of the geometric distribution used to describe it in the method are well known and can be calculated easily if infection rate β_{ij} for link between layers i and j is known

$$\mu_{ij} = 1/\beta_{ij} \tag{8}$$

$$\sigma_{ij}^2 = (1 - \beta_{ij})/\beta_{ij}^2 \tag{9}$$

which means in practice that for each layer l there is a different mean μ_l and variance σ_l^2, as well as $\mu_C = 1/\beta_C$ and $\sigma_C^2 = (1 - \beta_C)/\beta_C^2$ for interlayer links.

2.4. Evaluation metrics

Two efficiency measures are used for evaluating the quality of source detection: the average precision and the Credible Set Size at 0.95 confidence level. The precision for a single test is defined as the ratio between the number of correctly located sources and the number of sources found by the method. The tests are repeated multiple times for different sources and many graph realizations and then the obtained values of precision are averaged. The Credible Set Size at the confidence level of α (CSS_α) is the size of the smallest set of nodes containing the true source with probability α [51]. In other words this metric describes how many nodes with the highest *score* S_v should be labeled as the source to have the probability α that the true source is among these nodes.

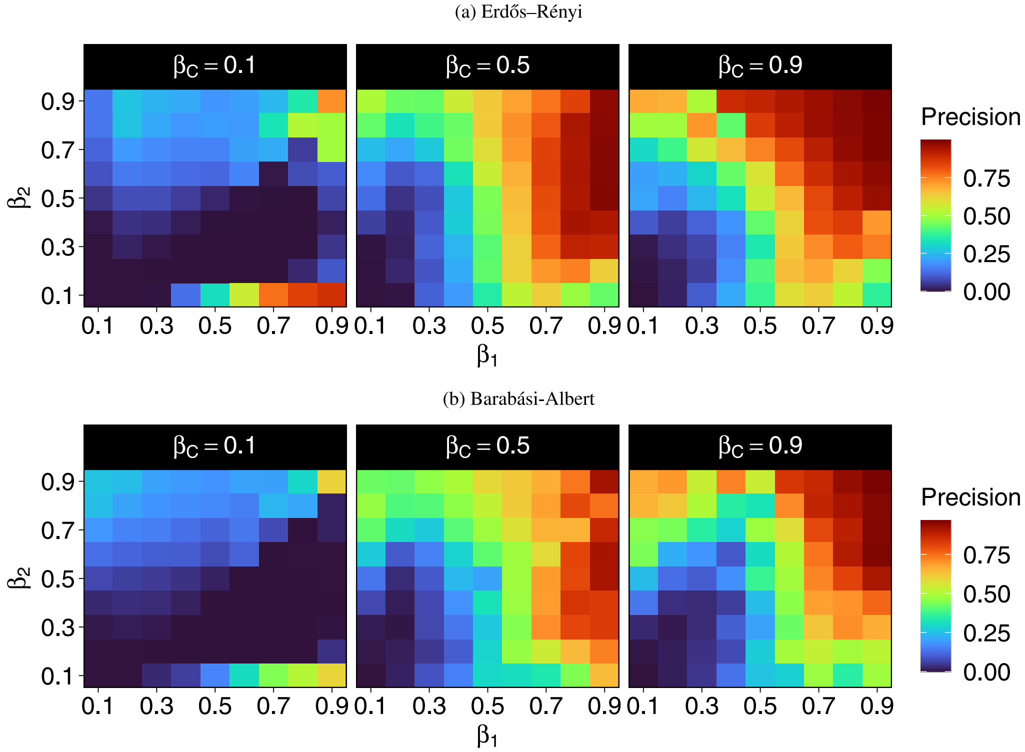

Fig. 3. Average precision of source localization in Erdős–Rényi (top) and Barabási–Albert (bottom) graphs with $L = 2$, $n_1 = n_2 = 500$ and $\langle k_1 \rangle = \langle k_2 \rangle = 8$. The observers are placed randomly with equal density in both layers $\rho_1 = \rho_2 = 0.1$. Layer 1 with spreading rate β_1 is a source layer. We consider three values of interlayer spreading rate β_C: 0.1 (left), 0.5 (middle) and 0.9 (right). The average precision is computed from 10^3 realizations.

3. Results

In this section we present the results of finding the source of an artificial SI spreading process on both synthetic and real multilayer networks. The first part concerns the case of two layers and shows the existence of two types of behavior – (i) *interference* and (ii) *synergy* between observers in the two network layers, depending mainly on how fast the spreading process between layers is. The following parts explore the case of multiple layers, comparison between synthetic and real multilayer networks as well as focused results of attempting to exploit multilayer structure to improve the quality of source localization.

3.1. Two layers

We study the performance of the source localization in bilayer Erdős–Rényi and Barabási–Albert networks for three values of the interlayer infection rate β_C: low (0.1), moderate (0.5) and high (0.9). The source of spreading is always placed in layer 1 which causes that the layers are distinguishable.

In the first experiment we vary the intralayer infection rates β_1, β_2 from 0.1 to 0.9. The observers are placed randomly with equal density in both layers $\rho_1 = \rho_2 = 0.1$. As can be seen in Fig. 3, the heat maps of the average precision for weak coupling between layers ($\beta_C = 0.1$) differ significantly from the heat maps for moderate and strong couplings (the heat maps for $CSS_{0.95}$ have the same properties thus they are not shown). In the former case, if $\beta_1 > 0.3$ then the average precision has two local maxima as function of β_2, one for $\beta_2 = 0.1$ and second for $\beta_2 = 0.9$. For higher values of the interlayer infection rate β_C (center and right panels in Fig. 3) the characteristics of the heat maps switches from bimodal to unimodal, with the only maximum for highest $\beta_2 = 0.9$). In that case the quality increases overall with the mean value of intralayer infection rate.

The asymmetry in bilayer network can be also caused by the difference of densities of observers within the layers, which is shown in Fig. 4. Here, the intralayer spreading rates are equal and moderate $\beta_1 = \beta_2 = 0.5$, but the density of observers ρ_1, ρ_2 vary from 0.02 to 0.2. As in the previous experiment, the source of spreading is always placed in layer 1. Again, the numerical simulations reveal substantial difference between the networks with weak and strong couplings, with Erdős–Rényi graphs changing behavior at lower β_C than Barabási–Albert networks. In the case of weak coupling between layers, the best quality of source location is achieved when the density of observers in the source layer is very

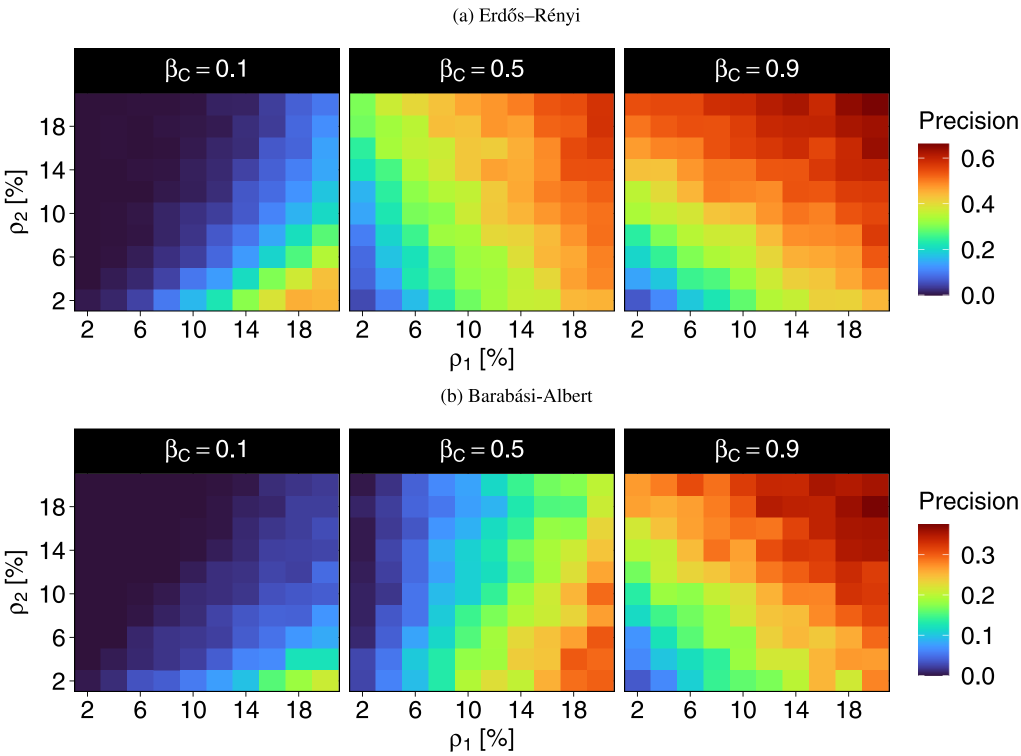

Fig. 4. Average precision of source localization in Erdős–Rényi (top) and Barabási–Albert (bottom) graphs with $L = 2$, $n_1 = n_2 = 500$ and $\langle k_1 \rangle = \langle k_2 \rangle = 8$. The intralayer spreading rates are $\beta_1 = 0.5$ (source layer) and $\beta_2 = 0.5$. We consider three values of interlayer spreading rate β_C: 0.1 (left), 0.5 (middle) and 0.9 (right). The average precision is computed from 10^3 realizations.

high $\rho_1 = 0.2$, and $\rho_2 = 0.02$ is very low. It means, that in this case, the additional observers placed in layer 2 not only do not help, but also *interfere* in the localization of the source in layer 1. This behavior changes when β_C is higher and the quality of source detection increases with the density of observers in any layer. It follows, there exists a certain critical value β_C^* that divides parameter space into two regimes.

3.2. More than two layers

We compare the performance of the source location algorithm for the networks with different number of layers according to two schemes. In the first one, all considered systems have the same number of nodes per layer n_l but they differ in total number of nodes n_{tot}, e.g. a four-layer graph has twice as many nodes as a two-layer one. In the second scheme, the total number of nodes n_{tot} is the same for all compared networks, which means that a two-layer graph has twice as many nodes per layer as a four-layer one. The results presented in Figs. 5 and 6 show strong influence of the number of layers on the average precision and $CSS_{0.95}$ for both the case of fixed n_l and fixed n_{tot}. We focus here on the case of $\beta_C > \beta_C^*$ and observe that increasing number of layers improves the source localization efficiency (higher precision, smaller $CSS_{0.95}$). A particularly large increase in the performance of the source location is observed after changing from one layer to two-layer graph. One can speculate that in the first scheme, the one with fixed number of nodes per layer n_l, the networks with larger number of layers have also more observers, but this is not the case in second scenario, when the number of observers depends only on n_{tot}, which on the other hand is independent of the number of layers L. This means that for high β_C not only observers in second layer do not interfere, but there is *synergy* that increases the accuracy of finding the source above the level of similar single layered network.

3.3. Experiments on the real-world multilayer network

As an example of real social network with many layers we use the data collected at the Department of Computer Science at Aarhus University among the employees [49]. The network is built of following layers:

- current working relationship,
- regularly eating lunch together,

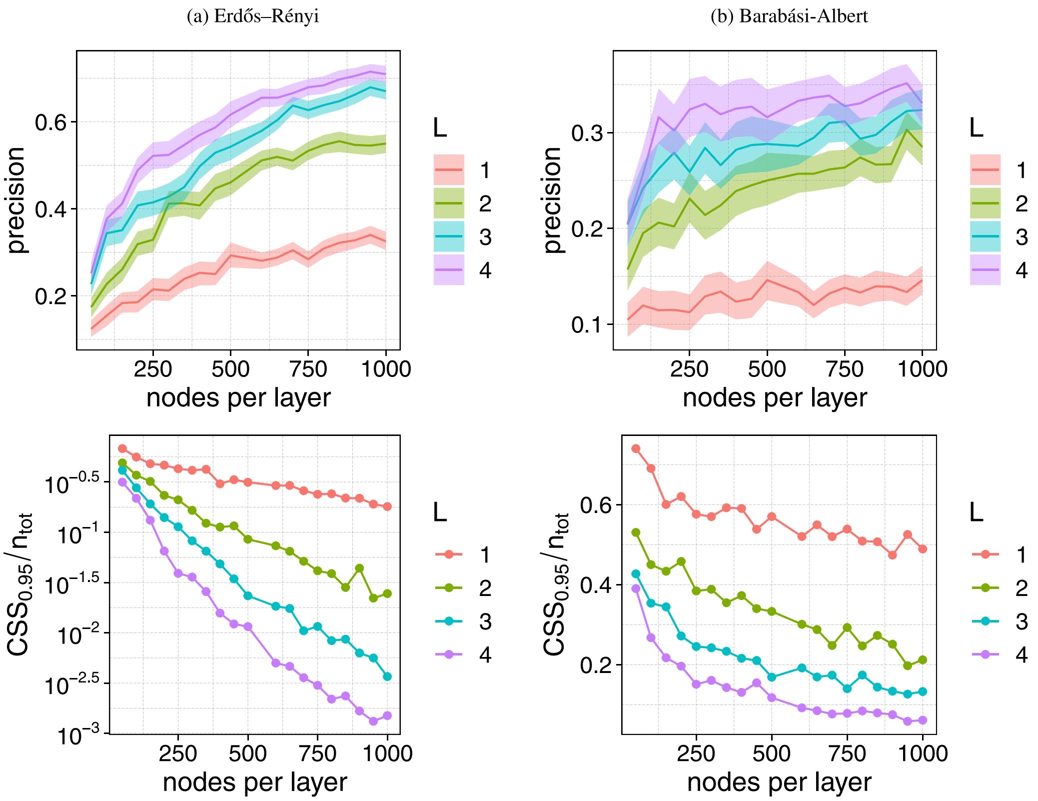

Fig. 5. The quality of source localization in Erdős–Rényi graph (left) and Barabási–Albert model (right) with different number of layers. The number of nodes per layer is the same for all networks, regardless of the number of layers L. The average degree $\langle k \rangle = 8$, density of observers $\rho = 0.1$ and intralayer infection rate $\beta = 0.5$ are the same in each layer. The interlayer infection rate is $\beta_C = 0.8$. The average precision is computed from 10^3 realizations. The ribbons show 95% Confidence Interval (CI).

Table 1
Basic properties of each layer of the Aarhus University network. The symbol n_c stands for the number of connected components, $\langle l \rangle$ for the average path length, Δ for the diameter and C for the global clustering coefficient.

Layer	$\lvert V \rvert$	$\langle k \rangle$	k_{max}	n_c	$\langle l \rangle$	Δ	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

- repeated leisure activities,
- friendship on Facebook.

Basic properties of each layer are shown in Table 1. We use the same spreading model and evaluation metrics for simulations as for aforementioned synthetic networks. In the first study with real topology, we choose two largest layers (Work & Lunch) and examine how the quality of source localization depends on the infection rates densities of observers in each layer. The results of these experiments shown in Fig. 7 are consistent with the corresponding heat maps of the average precision for the Erdős–Rényi and Barabási–Albert models which means our findings are not particular to the investigated synthetic topologies. Also the other pairs of layers (Work & Leisure, Work & Facebook, Work+Lunch & Facebook) were examined. These results are not shown here but they are coherent with the results for Work & Lunch layers and synthetic networks. Next, we explore the problem of many layers and investigate how the average precision depends on the coupling strength between the layers expressed by β_C. We do not limit our research to particular values of the infection rate β and density of observers ρ, but we perform the analysis for nine representative cases. The results presented in Fig. 8 show that for medium and high value of β (middle and bottom row) the multilayer structure of network can increase the efficiency of source identification if $\beta_C > \beta_C^*$. As seen in Fig. 8, the value of β_C^* is not constant but it depends on the intralayer infection rate β, density of observers ρ and even the number of layers L.

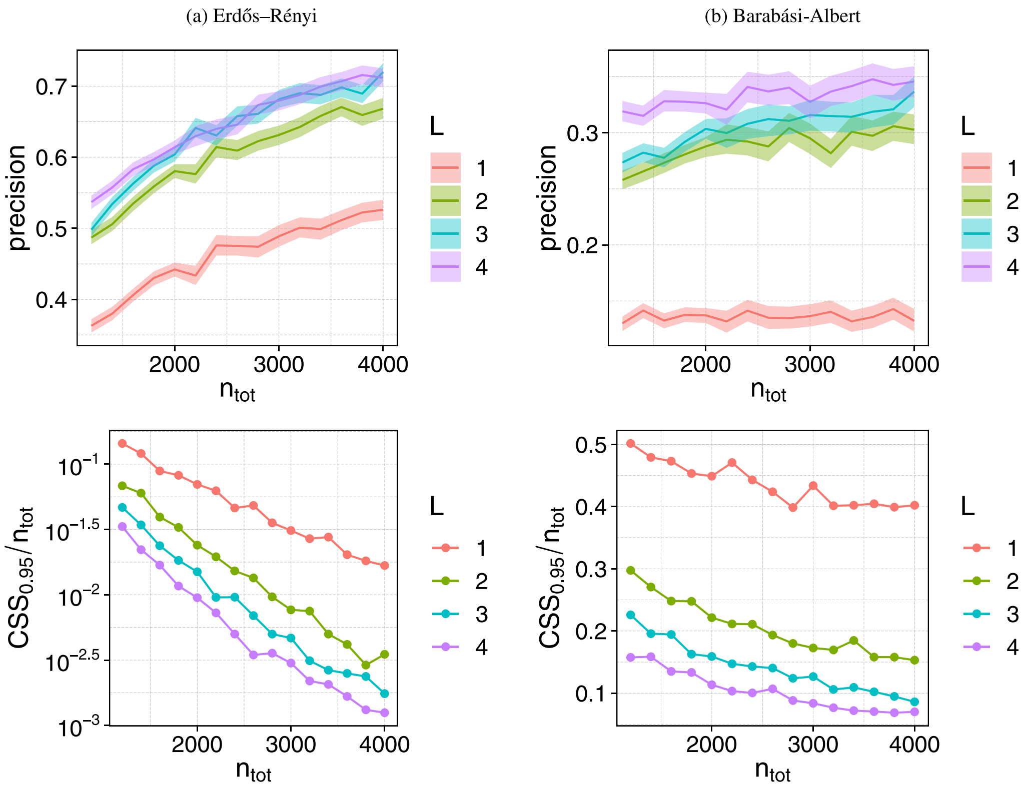

Fig. 6. The quality of source localization in Erdős–Rényi graph (left) and Barabási–Albert model (right) with different number of layers. The number of nodes per layer is inversely proportional to the number of layers L. The average degree $\langle k \rangle = 8$, density of observers $\rho = 0.1$ and intralayer infection rate $\beta = 0.5$ are the same in each layer. The interlayer infection rate is $\beta_C = 0.8$. The average precision is computed from 10^4 realizations. The ribbons show 95% Confidence Interval (CI).

3.4. Source finding for interfering layers

The results for both synthetic and real networks show (Fig. 4,7), that if the coupling between networks is low enough, then using observers in both layers yields worse results than just on the layer where the infection originated from. This suggests that it may be possible to improve the accuracy to find the source, if we can infer which layer the source is at, and only use observers in that layer to locate its position. We have studied the properties of the likelihood log-scores S_v (Eq. (6)) for both layers of 2-layer network and have found that the average log-scores are consistently higher for the layer that contained the true source, especially for the low-coupling regime (orange line on Fig. 9). In addition, the difference between average log-scores of nodes in both layers calculated using observers in one layer only (red line at Fig. 9) crosses or reaches 0 at β_C^* or in close proximity , where β_C^* is a critical value of β_C dividing the regions of interference and synergy. This value can be determined by looking at the point where the accuracy gain from taking observations from second layer (blue line at Fig. 9) crosses zero, which means we can tell whether the actual β_C is below or above critical value for the investigated system. After identifying the regime, if $\beta_C < \beta_C^*$ we can identify the layer where the true source is using average log-scores using all observers. Using that knowledge, we can discard the observers in layer not containing true source, improving accuracy and limiting the number of nodes we have to calculate log-score for (in case we used only the wrong layer observers to identify the regime in the first step). It is possible to calculate this without knowing true source, and hence in a practical situation where we need to locate the source.

This approach seems promising for investigated synthetic networks, but has not been thoroughly explored yet, as it may be possible that for some parameter ranges (network densities, specific topologies, spreading rate combinations in different layers) such simple method may not work correctly. More research on the topic is necessary, but the results provide a starting point for methods of exploiting multilayer structure of the network and improving the accuracy of finding the source via observer-based maximum likelihood methods.

4. Discussion

In this paper we have examined the issue of locating the source of Susceptible–Infected spreading process in multilayer networks using methods established for general networks [19] adapted to multilayer topology and exploring

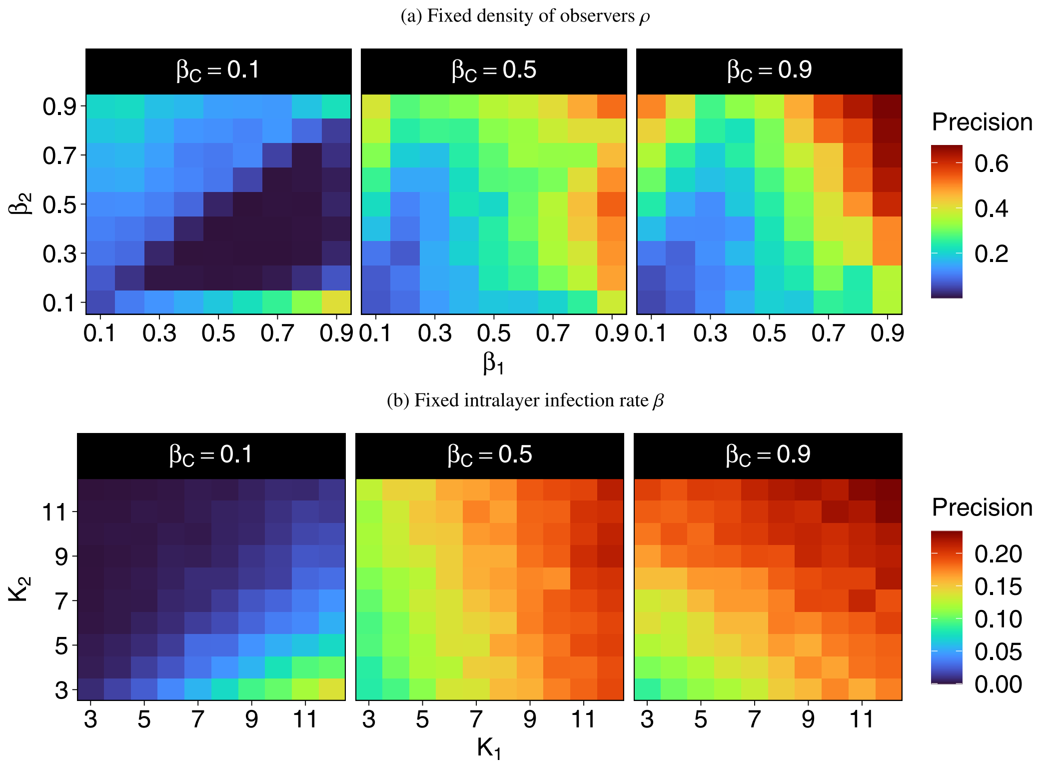

Fig. 7. Average precision of source localization in a bilayer graph built of two layers (Work, Lunch) from the Aarhus University network. We consider three values of interlayer spreading rate β_C: 0.1 (left), 0.5 (middle) and 0.9 (right). The true source is always in Layer 1. (a) The densities of observers are $\rho_1 = \rho_2 = 0.2$. (b) The intralayer spreading rates are $\beta_1 = \beta_2 = 0.5$. The average precision is computed from $5 \cdot 10^3$ realizations.

various network and spreading parameters. We have shown the existence of two different parameter ranges, one where observations in two layers interfere with each other lowering the accuracy of finding the source (as seen in Figs. 4 and 7), and the other where they synergize, allowing to find the true source easier (as seen in Fig. 6).

The first case happens for low spreading rate β_C between different layers. Here observers in the same layer as the true source prove to be useful and adding more of them increases accuracy, but observers in the other layer only lower it. If we use all observers then peak accuracy is achieved when one of the layers is dominant (has a much higher spreading rate), or when both have similar, high spreading rates i.e. the process is more deterministic. The threshold for β_C is not fixed and depends on the topology among other factors — scale-free networks have higher thresholds than simple random graphs. We believe the interference from second layer observers to be caused by misalignment between assumptions in the method (tree graph) and real structure (network with loops) [25], similar to the loss of precision from faraway observers in scale-free graphs [22].

The second case happens when interlayer spreading rate β_C is high. It is much more intuitive, as additional observers increase the accuracy as would be expected from additional information. The nontrivial result here is that the increased accuracy ends up higher than for a single layer network of the same size and observer density. It means that a multilayer structure, given high enough interlayer coupling, is more conductive to finding the source than single layer topologies.

In addition, we have formulated a simple heuristic method to increase the accuracy of locating the source in multilayer networks that can identify whether the observers in different layers interfere or synergize and point to the layer containing the true source. This allows us to discard the second layer if we find ourselves with the interfering observers case. The method relies on mean likelihood scores that can be always calculated and thus can be always applied if dealing with a multilayer network.

It is worth noting that the SI model can be treated as special case of Susceptible–Infected–Removed (SIR) or rumor model [52,53], where agents do not remain active forever (setting the recovery rate $\gamma = 0$ in those models produces an SI process). The location method relies on first arrival times however, meaning that unless the deactivation is very strong (like in independent cascade model where $\gamma = 1$) such that it would significantly alter the characteristics of the delay distribution or limit the number of observers that receive the information at all, its impact on the accuracy should be small. This is especially true given the already present mismatch between the distribution assumed by source location method and the actual delay distribution in the SI model.

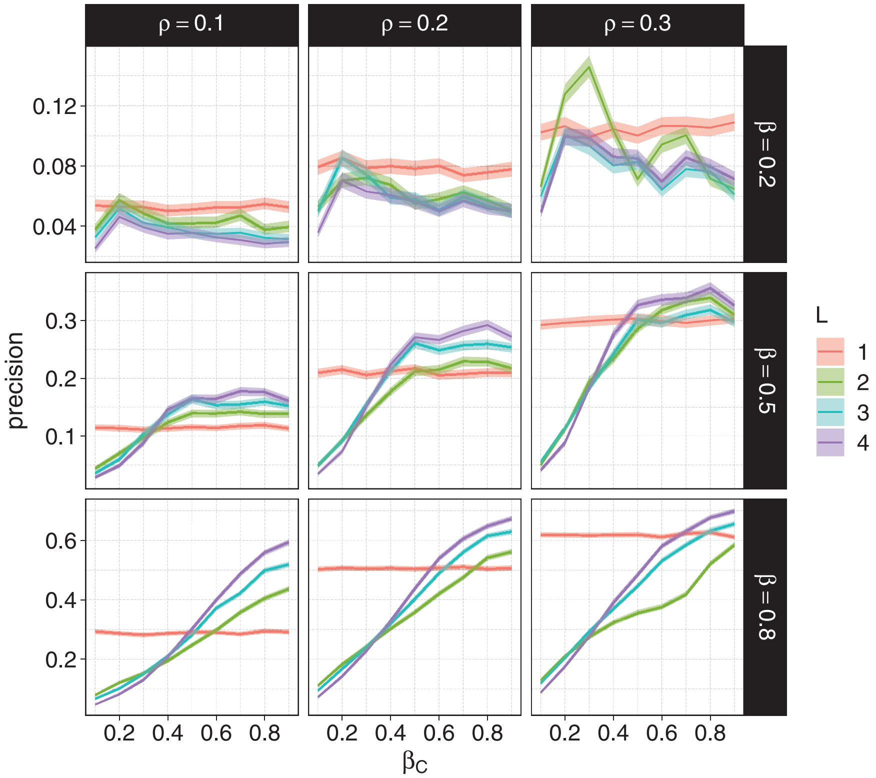

Fig. 8. Average precision of source localization in the Aarhus University network as a function of interlayer spreading rate β_C for different number of layers L considered in spreading and backtracing. The layers are added from largest (Work) to smallest (Facebook). We consider three values of density of observers ρ: 0.1 (left), 0.2 (center), 0.3 (right) and also three values of the intralayer spreading rate β (which is the same for each layer): 0.2 (top), 0.5 (middle), 0.8 (bottom). The average precision is computed from 10^4 realizations.

(a) Erdős–Rényi (b) Barabási-Albert

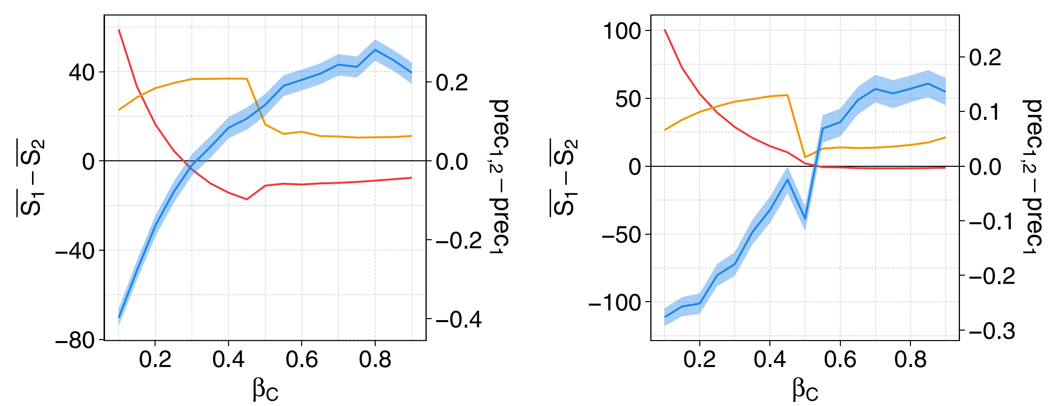

Fig. 9. Difference between average likelihood log-scores S_v for nodes in layer 1 (where the true source is located) and layer 2 of a 2-layer network (orange line), depending on the interlayer spreading rate β_C as well as difference between average log-scores of layers calculated using only observers in one layer (red line). The difference in precision between using observers in both layers ($\text{prec}_{1,2}$) and only layer 1 (prec_1) is also shown for comparison (blue line with 95% CI). The results are for networks, of $n_1 = n_2 = 500$ nodes and mean degree $\langle k_1 \rangle = \langle k_2 \rangle = 8$, with intralayer spreading $\beta_1 = \beta_2 = 0.5$ and $5 \cdot 10^3$ realizations. It is important to note that the separation between average log-scores holds even for a single realization, not only for an aggregate over realizations shown in the figure.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The project supporting Ł.G.G., K.S. and J.A.H was funded by POB Research Centre Cybersecurity and Data Science of Warsaw University of Technology, Poland within the Excellence Initiative Program — Research University (ID-UB). R.P. was partially supported by the National Science Centre, Poland, agreement No. 2019/32/T/ST6/00173, and by PLGrid Infrastructure. J.A.H. was partially supported by the Russian Science Foundation, Agreement No. 17-71-30029 with co-financing of Bank Saint Petersburg, Russia.

References

[1] D. Stauffer, World records in the size of simulated Ising models, Braz. J. Phys. 30 (4) (2000) 787–793, http://dx.doi.org/10.1590/S0103-97332000000400029.
[2] K. Kacperski, J. Hołyst, Phase transitions and hysteresis in a cellular automata-based model of opinion formation, J. Stat. Phys. 84 (1–2) (1996) 169–189, http://dx.doi.org/10.1007/BF02179581.
[3] J. Holyst, K. Kacperski, F. Schweitzer, Phase transitions in social impact models of opinion formation, Physica A 285 (1) (2000) 199–210, http://dx.doi.org/10.1016/S0378-4371(00)00282-X.
[4] J. Holyst, K. Kacperski, F. Schweitzer, Social impact models of opinion dynamics, Annu. Rev. Comput. Phys. IX (2001) 253–273, http://dx.doi.org/10.1142/9789812811578_0005.
[5] A. Aleksiejuk, J. Hołyst, D. Stauffer, Ferromagnetic phase transition in Barabási-Albert networks, Physica A 310 (1–2) (2002) 260–266, http://dx.doi.org/10.1016/S0378-4371(02)00740-9.
[6] A. Aleksiejuk, J. Hołyst, A simple model of bank bankruptcies, Physica A 299 (1–2) (2001) 198–204, http://dx.doi.org/10.1016/S0378-4371(01)00296-5.
[7] A. Aleksiejuk, J. Hołyst, G. Kossinets, Self-organized criticality in a model of collective bank bankruptcies, Internat. J. Modern Phys. C 13 (3) (2002) 333–341, http://dx.doi.org/10.1142/S0129183102003164.
[8] D. Stauffer, J. Hołyst, Transfer in multi-theme opinion dynamics of Deffuant et al, 2008, arXiv:0803.2025.
[9] D. Stauffer, P.A. Grabowicz, J.A. Holyst, Opinion dynamics with hopfield neural networks, 2008, arXiv:0712.4364.
[10] D. Shah, T. Zaman, Rumors in a network: Who's the culprit? IEEE Trans. Inform. Theory 57 (8) (2011) 5163–5181, http://dx.doi.org/10.1109/TIT.2011.2158885.
[11] B.A. Prakash, J. Vrekeen, C. Faloutsos, Spotting culprits in epidemics: How many and which ones? in: Proceedings - IEEE International Conference on Data Mining, ICDM, 2012, pp. 11–20, http://dx.doi.org/10.1109/ICDM.2012.136.
[12] W. Luo, W.P. Tay, Estimating infection sources in a network with incomplete observations, in: 2013 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2013 - Proceedings, 2013, pp. 301–304, http://dx.doi.org/10.1109/GlobalSIP.2013.6736875.
[13] J. Jiang, S. Wen, S. Yu, Y. Xiang, W. Zhou, K-center: An approach on the multi-source identification of information diffusion, IEEE Trans. Inf. Forensics Secur. 10 (12) (2015) 2616–2626, http://dx.doi.org/10.1109/TIFS.2015.2469256.
[14] D. Brockmann, D. Helbing, The hidden geometry of complex, network-driven contagion phenomena, Science 342 (6164) (2013) 1337–1342, http://dx.doi.org/10.1126/science.1245200.
[15] A.Y. Lokhov, M. Mézard, H. Ohta, L. Zdeborová, Inferring the origin of an epidemic with a dynamic message-passing algorithm, Phys. Rev. E 90 (1) (2014) 1–9, http://dx.doi.org/10.1103/PhysRevE.90.012801.
[16] K. Zhu, L. Ying, Information source detection in the SIR model: A sample-path-based approach, IEEE/ACM Trans. Netw. 24 (1) (2016) 408–421, http://dx.doi.org/10.1109/TNET.2014.2364972.
[17] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, R. Zecchina, Bayesian Inference of epidemics on networks via belief propagation, Phys. Rev. Lett. 112 (11) (2014) 1–5, http://dx.doi.org/10.1103/PhysRevLett.112.118701.
[18] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Ingrosso, R. Zecchina, The patient-zero problem with noisy observations, J. Stat. Mech. Theory Exp. 2014 (10) (2014) P10016, http://dx.doi.org/10.1088/1742-5468/2014/10/P10016.
[19] P. Pinto, P. Thiran, M. Vetterli, Locating the source of diffusion in large-scale networks, Phys. Rev. Lett. 109 (6) (2012) http://dx.doi.org/10.1103/PhysRevLett.109.068702.
[20] N. Karamchandani, M. Franceschetti, Rumor source detection under probabilistic sampling, in: IEEE International Symposium on Information Theory - Proceedings, 2013, pp. 2184–2188, http://dx.doi.org/10.1109/ISIT.2013.6620613.
[21] W. Luo, W.P. Tay, M. Leng, How to identify an infection source with limited observations, IEEE J. Sel. Top. Sign. Proces. 8 (4) (2014) 586–597, http://dx.doi.org/10.1109/JSTSP.2014.2315533.
[22] R. Paluch, X. Lu, K. Suchecki, B.K. Szymański, J.A. Hołyst, Fast and accurate detection of spread source in large complex networks, Sci. Rep. 8 (1) (2018) 2508, http://dx.doi.org/10.1038/s41598-018-20546-3.
[23] M. Fang, P. Shi, W. Shang, X. Yu, T. Wu, Y. Liu, Locating the source of asynchronous diffusion process in online social networks, IEEE Access 6 (2018) 17699–17710, http://dx.doi.org/10.1109/ACCESS.2018.2817553.
[24] X. Li, X. Wang, C. Zhao, X. Zhang, D. Yi, Locating the source of diffusion in complex networks via Gaussian-based localization and deduction, Appl. Sci. 9 (18) (2019) http://dx.doi.org/10.3390/app9183758.
[25] Ł. Gajewski, K. Suchecki, J. Hołyst, Multiple propagation paths enhance locating the source of diffusion in complex networks, Physica A 519 (2019) 34–41, http://dx.doi.org/10.1016/j.physa.2018.12.012.
[26] S. Xu, C. Teng, Y. Zhou, J. Peng, Y. Zhang, Z.K. Zhang, Identifying the diffusion source in complex networks with limited observers, Physica A 527 (2019) 121267, http://dx.doi.org/10.1016/j.physa.2019.121267.
[27] H.-J. Wang, K.-J. Sun, Locating source of heterogeneous propagation model by universal algorithm, Europhys. Lett. 131 (4) (2020) http://dx.doi.org/10.1209/0295-5075/131/48001.
[28] X. She, X. Li, Y. Liu, C. Gao, A novel source locating strategy without consistent assumptions, in: 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD, 2016, pp. 702–708, http://dx.doi.org/10.1109/FSKD.2016.7603260.
[29] Q. Huang, C. Zhao, X. Zhang, D. Yi, Locating the source of spreading in temporal networks, Physica A 468 (2017) 434–444, http://dx.doi.org/10.1016/j.physa.2016.10.081.
[30] J. Jiang, S. Wen, S. Yu, Y. Xiang, W. Zhou, Rumor source identification in social networks with time-varying topology, IEEE Trans. Dependable Secure Comput. 15 (1) (2018) 166–179, http://dx.doi.org/10.1109/TDSC.2016.2522436.
[31] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M.A. Porter, S. Gómez, A. Arenas, Mathematical formulation of multilayer networks, Phys. Rev. X 3 (4) (2014) 1–15, http://dx.doi.org/10.1103/PhysRevX.3.041022.
[32] M. De Domenico, V. Nicosia, A. Arenas, V. Latora, Structural reducibility of multilayer networks, Nature Commun. 6 (2015) 1–9, http://dx.doi.org/10.1038/ncomms7864.

[33] M. Kivelä, A. Arenas, M. Barthelemy, J. Gleeson, Y. Moreno, M. Porter, Multilayer networks, J. Complex Netw. 2 (3) (2014) 203–271, http://dx.doi.org/10.1093/comnet/cnu016.
[34] S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, Phys. Rep. 544 (1) (2014) 1–122, http://dx.doi.org/10.1016/j.physrep.2014.07.001.
[35] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, A. Arenas, Ranking in interconnected multilayer networks reveals versatile nodes, Nature Commun. 6 (1) (2015) 1–6, http://dx.doi.org/10.1038/ncomms7868.
[36] A. Aleta, Y. Moreno, Multilayer networks in a nutshell, Annu. Rev. Condens. Matter Phys. 10 (1) (2019) 45–62, http://dx.doi.org/10.1146/annurev-conmatphys-031218-013259.
[37] G. Menichetti, D. Remondini, P. Panzarasa, R.J. Mondragón, G. Bianconi, Weighted multiplex networks, PLoS One 9 (6) (2014) 6–13, http://dx.doi.org/10.1371/journal.pone.0097857.
[38] Ł.G. Gajewski, J. Sienkiewicz, J.A. Hołyst, Discovering hidden layers in quantum graphs, 2021, arXiv:2012.01454.
[39] Ł.G. Gajewski, J. Sienkiewicz, J.A. Hołyst, Bifurcations and catastrophes in temporal bi-layer model of echo chambers and polarisation, 2021, arXiv:2101.03430.
[40] Ł.G. Gajewski, J. Chołoniewski, M. Wilinski, Detecting hidden layers from spreading dynamics on complex networks, 2021, arXiv:2101.11758.
[41] A. Barrat, M. Barthelemy, A. Vespignani, Dynamical Processes on Complex Networks, Cambridge university press, 2008, http://dx.doi.org/10.1017/CBO9780511791383.
[42] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Modern Phys. 87 (3) (2015) 925, http://dx.doi.org/10.1103/RevModPhys.87.925.
[43] M. De Domenico, C. Granell, M.A. Porter, A. Arenas, The physics of spreading processes in multilayer networks, Nat. Phys. 12 (10) (2016) 901–906, http://dx.doi.org/10.1038/nphys3865.
[44] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C.J. Perez-Vicente, Y. Moreno, A. Arenas, Diffusion dynamics on multiplex networks, Phys. Rev. Lett. 110 (2) (2013) 028701, http://dx.doi.org/10.1103/PhysRevLett.110.028701.
[45] A. Sole-Ribalta, M. De Domenico, N.E. Kouvaris, A. Diaz-Guilera, S. Gomez, A. Arenas, Spectral properties of the Laplacian of multiplex networks, Phys. Rev. E 88 (3) (2013) 032807, http://dx.doi.org/10.1103/PhysRevE.88.032807.
[46] G.F. de Arruda, E. Cozzo, T.P. Peixoto, F.A. Rodrigues, Y. Moreno, Disease localization in multilayer networks, Phys. Rev. X 7 (1) (2017) 1–25, http://dx.doi.org/10.1103/PhysRevX.7.011014.
[47] G.F. de Arruda, F.A. Rodrigues, Y. Moreno, Fundamentals of spreading processes in single and multilayer complex networks, Phys. Rep. 756 (2018) 1–59, http://dx.doi.org/10.1016/j.physrep.2018.06.007.
[48] E. Cozzo, R.A. Baños, S. Meloni, Y. Moreno, Contact-based social contagion in multiplex networks, Phys. Rev. E 88 (5) (2013) 1–5, http://dx.doi.org/10.1103/PhysRevE.88.050801.
[49] M. Magnani, B. Micenkova, L. Rossi, Combinatorial analysis of multiple networks, 2013, arXiv:1303.4986.
[50] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. A 115 (772) (1927) 700–721, http://dx.doi.org/10.1098/rspa.1927.0118.
[51] R. Paluch, Ł.G. Gajewski, J.A. Hołyst, B.K. Szymanski, Optimizing sensors placement in complex networks for localization of hidden signal source: A review, Future Gener. Comput. Syst. 112 (2020) 1070–1092, http://dx.doi.org/10.1016/j.future.2020.06.023.
[52] D.J. Daley, D.G. Kendall, Epidemics and rumors, Nature 204 (1964) 1964, http://dx.doi.org/10.1038/2041118a0.
[53] M. Nekovee, Y. Moreno, G. Bianconi, M. Marsili, Theory of rumour spreading in complex social networks, Physica A 374 (1) (2007) 457–470, http://dx.doi.org/10.1016/j.physa.2006.07.017.