



# Multiple propagation paths enhance locating the source of diffusion in complex networks

Ł.G. Gajewski<sup>a,\*</sup>, K. Suchecki<sup>a</sup>, J.A. Hołyst<sup>a,b</sup>

<sup>a</sup> Center of Excellence for Complex Systems Research, Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00662 Warsaw, Poland

<sup>b</sup> ITMO University, 49 Kronverkskiy av., 197101 Saint Petersburg, Russia

## HIGHLIGHTS

- Multiple paths reduce time it takes to spread.
- We find approximations for mean times and covariances as if paths or their links were equally probable.
- These approximations result in 1.6 times higher accuracy than single path approximation.

## ARTICLE INFO

### Article history:

Received 24 August 2018

Available online 19 December 2018

### Keywords:

Complex networks

Spreading

Diffusion

Source location

Maximum likelihood

## ABSTRACT

We investigate the problem of locating the source of diffusion in complex networks without complete knowledge of nodes' states. Some currently known methods assume the information travels via a single, shortest path, which by assumption is the fastest way. We show that such a method leads to the overestimation of propagation time for synthetic and real networks, where multiple shortest paths as well as longer paths between vertices exist. We propose a new method of source estimation based on maximum likelihood principle, that takes into account existence multiple shortest paths. It shows up to 1.6 times higher accuracy in synthetic and real networks.

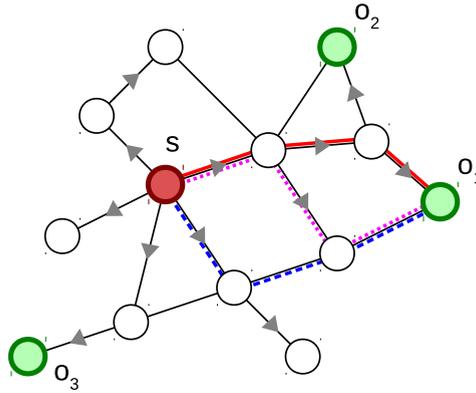
© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Understanding how information propagates in a system is an important field of study in complex networks. The information can be of a various nature - e.g. it could be a virus [1] or a tweet [2]. The main branch of studying epidemics and diffusion is about modelling the propagation itself [3–5]. We consider a reverse problem – locating the source, i.e. “the patient zero”. That reverse problem has already been approached before. Shah and Zaman introduced a measure of *rumour centrality* defined as the number of distinct ways a rumour can spread from a given node, thus providing with a maximum likelihood estimator of a rumour's source. That method, however, requires the full knowledge of both topology of a network and states of all vertices [6]. Pinto, Thiran, Vetterli (PTV), on the other hand, while still requiring full knowledge of network's topology, provide a solution with information about the states of only some of the nodes needed [7]. An alternative approach to PTV that retains the limited knowledge assumptions has been introduced by Shen et al. which provides the source estimator via minimizing the variance of a time vector of a backwards spread signal [8]. More recent research shows that there are possible solutions for locating a source in temporal networks [9] and that an efficient placement of observers is not a trivial task because different common strategies do not significantly differ in resulting accuracy [10]. The aforementioned

\* Corresponding author.

E-mail address: [lukaszgajewski@tuta.io](mailto:lukaszgajewski@tuta.io) (Ł.G. Gajewski).



**Fig. 1.** A schematic illustration of the problem. Vertex  $S$  is the source we are looking for. Nodes  $O_i$  are our observers. One can clearly see that there are three shortest topological paths between the source and vertex  $O_1$ . Moreover those paths are correlated, i.e. there are edges that are common for several shortest paths between  $S$  and  $O_1$ . What is more, since delays on edges are random variables a shorter path is not always faster, e.g. in the illustration the shortest path between  $S$  and  $O_2$  is not the fastest one and so the signal travels via a topologically longer route.

PTV method has also been recently improved by Paluch et al. in both accuracy (for scale-free networks) and computation time by limiting observers to closest ones instead of using all, and using gradient to select nodes that have likelihood calculated at all [11].

## 2. Set up

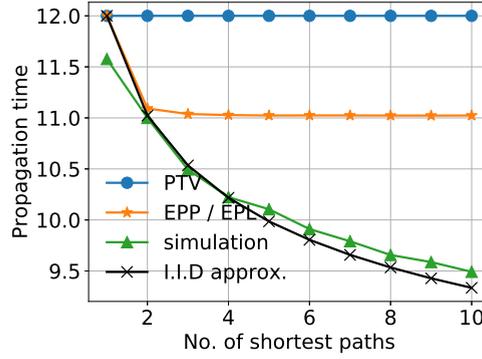
Consider an arbitrary graph that can represent cities connected via highways or friends on Facebook etc. One of the nodes sends some information or a signal to its neighbours. Then those newly “infected” vertices keep passing the message on until all nodes receive it. We can either assume that the signal is transmitted to node’s neighbours with a certain probability each time step (SI model) or that it is always transmitted with a random delay from an arbitrary distribution (Gaussian in our study). We assume that we know the topology of the network as well as the distribution from which the delays on the links are sampled. Additionally some subset of all vertices provide us information about at what time they received the signal (we shall call those nodes *observers*). Our goal is to locate the source of the signal (i.e. the node that generated it).

PTV algorithm assumes that information spreads through shortest paths and while it is intuitively satisfying it is also just an approximation. In the actual process, there may exist multiple paths of propagation, not only a single one (see Fig. 1). Assuming that on each edge we have an unknown delay from a known probability distribution and that this distribution is Gaussian (as assumed by Pinto et al.) we can compare PTV’s expected time of information’s traversal between two nodes to the simulated one. Such comparison is shown at Fig. 2. There are two reasons for the discrepancy between PTV and the simulation. First, there can be more than one shortest path (and usually this is the case in non-tree networks). Since the information travels via *fastest* path, every additional parallel path gives the process another chance for the information to arrive quickly. Instead of distribution of single path traversal time, we have distribution of minimum time from among several different path times. Even if mean for each path is the same, the minimum among them has a different distribution than a single path, with a smaller mean value. Second reason is the existence of independent longer paths, that may end up faster than all shortest paths by chance. Thus existence of these additional chances for quick traversal (even if much less probable than shortest paths) still decrease mean time. If the variance of the times compared to mean is small, contribution of longer paths is negligible, but increases as variance becomes larger. We have focused on the first of the two reasons of discrepancy and developed a new method based on maximum likelihood, centred around existence of multiple shortest paths. Similar to PTV, we assume normal distributions of delays on edges, and assume that resulting distribution of arrival at different observers is still multivariate normal distribution, or at least it can be approximated as such. We have therefore to find (i) the mean times of traversal between source and all observers, (ii) the covariance matrix of these times.

Let us assume we have a general graph and each edge has a delay sampled from a known Gaussian distribution  $N(\mu, \sigma^2)$ . Particular delays shall remain unknown to the source locating algorithm. Some randomly chosen subset of nodes (observers) register time of reception as sum of weights along the fastest path between source of information and node in question. The maximum likelihood for multivariate normal distribution of times would be:

$$\hat{s} = \operatorname{argmax}_{s \in G} \frac{1}{\sqrt{\mathcal{T}^n} \sqrt{|\Sigma_s|}} \exp\left(-\frac{1}{2}(\mathbf{d} - \boldsymbol{\mu}_s)^T \Sigma_s^{-1} (\mathbf{d} - \boldsymbol{\mu}_s)\right) \quad (1)$$

where  $\mathbf{d}$  is observed delay vector (each element corresponds to a time at which given observer received the information relatively to the reference observer),  $\mathcal{T} = 2\pi$ ,  $\boldsymbol{\mu}_s$  is a deterministic delay vector (each element is an expected time of transmission relative to reference observer) if  $s$  is the source,  $\Sigma_s$  is a covariance matrix assuming source  $s$ ,  $\hat{s}$  is our estimated source and  $G$  is the whole graph. PTV assumes information always spreads along single set of shortest paths, that form a



**Fig. 2.** Signal's propagation time between two nodes in a network as a function of number of shortest paths connecting those nodes. Simulations were conducted on Barabási–Albert graph with 100 nodes, with propagation ratio  $\frac{\mu}{\sigma} = 4$ , two nodes were chosen so that their shortest paths length is set to three edges. Line labelled *simulation* is the real, i.e. experimental time whereas *PTV* is time assumed in that method (shortest path length times  $\mu$ ), *EPP/EPL* is result of taking into account up to two shortest paths (that is what we use for our modifications) see Eq. (13) and *I.I.D* is what we receive when taking all shortest paths into consideration yet assuming their uncorrelated (do not have common edges).

breadth-first search tree (*BFS*) rooted at a candidate node  $s$ . Deterministic delays  $\mu_s$  in such case are equal to  $\mu * L_s$ , where  $L_s$  are lengths of the shortest (and the only existing) paths between source  $s$  and observers. Covariance matrix elements are intersections of shortest path between two given observers to the reference observer (and simply lengths of paths on the diagonal) multiplied by  $\sigma^2$ . We shall reject the *BFS* approach and deal with whole sets of shortest paths, evaluating  $\mu_s$  and  $\Sigma$  differently.

### 3. Independent paths

Let all paths between two nodes be represented by Gaussian random variables  $X_1, X_2, \dots, X_n$  with known joint probability distribution

$$f(\mathbf{X}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right)}{\sqrt{|\boldsymbol{\Sigma}|}} \quad (2)$$

Where  $\mathbf{X} = [X_1, X_2, \dots, X_n]$ ,  $\boldsymbol{\mu}$  is mean values vector and  $\boldsymbol{\Sigma}$  is a covariance matrix. We would love to have an expression for an expected value of propagation time -  $\mathbf{E}[X_{min}] = \mathbf{E}[\min(\mathbf{X})] = \mathbf{E}[\min(X_1, X_2, \dots, X_n)]$ . The probability density function is given by

$$\begin{aligned} \phi(X_{min}) &= -\frac{d}{dX_{min}} P(X_1, X_2, \dots, X_n > X_{min}) = \\ &= -\frac{d}{dX_{min}} \int_{X_{min}}^{\infty} \dots \int_{X_{min}}^{\infty} f(\mathbf{X}) dX_1 dX_2 \dots dX_n \end{aligned} \quad (3)$$

then the expected value  $X_{min}$

$$\mathbf{E}[X_{min}] = \int_{-\infty}^{\infty} x \phi(x) dx \quad (4)$$

One can clearly see this is far from a trivial task. We are forced to make some simplifications. One possible approach is to assume there are no correlations amongst  $X_i$ , i.e. we deal with independently, identically distributed variables (I.I.D):

$$\Phi_{X_{min}}(x) = P(X_{min} \leq x) = 1 - P(\min(X_1, X_2, \dots, X_n) > x) \quad (5)$$

where  $\Phi_{X_{min}}(x)$  is the cumulative distribution function of  $X_{min}$ , also

$$\min(X_1, X_2, \dots, X_n) > x \iff \forall_{i=1,2,\dots,n} X_i > x \quad (6)$$

then

$$\begin{aligned} \Phi_{X_{min}}(x) &= 1 - P(X_1 > x)P(X_2 > x) \dots P(X_n > x) = 1 - P(X_1 > x)^n = \\ &= 1 - [1 - P(X_1 \leq x)]^n = 1 - [1 - \Phi_{X_1}(x)]^n \end{aligned} \quad (7)$$

where  $\Phi_{X_1}$  is given by

$$\Phi_{X_1}(x) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{x - \mu}{\sqrt{2}\sigma} \right) \right) \quad (8)$$

The probability density function of  $X_{min}$  is then given by

$$\phi_{X_{min}}(x) = \frac{d\Phi_{X_{min}}(x)}{dx} \tag{9}$$

After some working out we conclude that

$$\begin{aligned} E[X_{min}] &= \int_{-\infty}^{\infty} x\phi_{X_{min}}(x)dx = \\ &= \int_{-\infty}^{\infty} \frac{xn \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{T}\sigma} \left[1 - \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right)\right]^{n-1} dx \end{aligned} \tag{10}$$

Alternatively we can also take into account that negative delays are non-existent in vast majority of real world scenarios (i.e.  $x \geq 0$ ) and then we have

$$\Phi_{X_1}(x) = \frac{\operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)}{1 + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)} \tag{11}$$

$$E[X_{min}] = \int_{-\infty}^{\infty} \frac{xn \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{T}\sigma} \left[1 - \frac{\operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)}{1 + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)}\right]^{n-1} dx \tag{12}$$

Ignoring that, however, can be justified if *propagation ratio*  $-\frac{\mu}{\sigma}$  is sufficiently large (due to the nature of *error function*).

Comparison of the above approximation (10) with simulation, PTV and alternative approach chosen for our modifications is shown at Fig. 2. Results are for propagation ratio  $\frac{\mu}{\sigma} = 4$  and one could say they are promising, however, for large values of  $\sigma$  or for strongly correlated paths the results are less satisfying. While that provides a vast improvement over original PTV approach the nature of the integral is such that it takes substantially more time to calculate and so its benefits might not necessarily exceed its costs.

#### 4. Multiple correlated paths

Since establishing general expression for the expected value of minimum distribution of an ensemble of  $n$  shortest paths is a non-trivial task, even with I.I.D approximation described above, we shall settle for a known analytical formula in the case of  $n = 2$  derived in [12].

Let  $Y = \min(X_1, X_2)$  where  $X_1, X_2$  are known Gaussian distributions such that  $E[X_i] = \mu_i, E[X_i^2] - E[X_i]^2 = \sigma_i$ , and  $\Phi, \phi$  are respectively the cumulative distribution function and probability d.f. of the standard normal distribution. Then:

$$E[Y] = \mu_1\Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + \mu_2\Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) - \theta\phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) \tag{13}$$

$$E[Y^2] = (\sigma_1^2 + \mu_1^2)\Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + (\sigma_2^2 + \mu_2^2)\Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) - (\mu_1 + \mu_2)\theta\phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) \tag{14}$$

$$\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{15}$$

Where  $\rho$  is a correlation coefficient between the distributions  $X_1, X_2$  that in case of paths we define as number of common edges between the paths  $R_1, R_2$  normalized by the length  $L$  of the path:

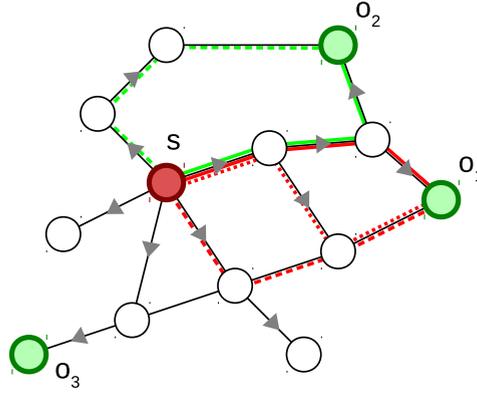
$$\rho = \frac{R_1 \cap R_2}{L} \tag{16}$$

Now evaluate elements of  $\mu$  using (13) if  $n > 1$  else as in PTV. In case there are more than two shortest paths apply (13) for two least correlated ones (that is those who have the smallest number of common edges). Result of such approach compared with PTV and “real” times in simulation is presented at Fig. 2. As one can clearly see it is still far from truth yet significantly closer than PTV.

##### 4.1. Equiprobable paths (EPP)

Since we have rejected *BFS* trees we cannot calculate covariance matrices as before. A natural generalization of PTV’s formula is to use expected value of covariance between all shortest paths from node  $o_{i+1}$  to reference observer  $o_1$  and all shortest paths between  $o_{j+1}$  and  $o_1$  while on diagonal elements instead of  $\sigma^2 * L$  we calculate the variance of minimum distribution if  $n > 1$  (Fig. 3). I.e.: Let  $\heartsuit$  represent a set of paths each treated as random variable:

$$\heartsuit_{i+1} := R_1(o_1, o_{i+1}), R_2(o_1, o_{i+1}), \dots, R_n(o_1, o_{i+1}) \tag{17}$$



**Fig. 3.** Correlations between paths between supposed source  $S$  and observers  $O_1$  (red paths) and  $O_2$  (green paths). Approach EPP assumes all paths are equally probable and calculates expected overlap between all pairs of red–green paths (out of six pairs, one has overlap 1, one has overlap 2). Approach EPL assumes all links are equally probable to be used by real spreading and calculates overlap between union of all red and union of all green paths (two links), normalized by union of all paths (eleven links). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where  $R_i(x, y)$  is an  $i$  –  $th$  path between vertices  $x$  and  $y$ , then:

$$\Sigma_{i,j} = \begin{cases} \sigma^2 \sum_{k=1}^n \sum_{l=1}^m p_{k,l} \cdot |R_k(o_1, o_{i+1}) \cap R_l(o_1, o_{j+1})| & i \neq j \\ \min_{\sigma^2}(\heartsuit_{i+1}) = E[Y^2] - E[Y]^2 & i = j \end{cases} \quad (18)$$

Where  $p_{k,l}$  is a probability that  $k$  –  $th$  path from the set of paths connecting nodes  $o_1, o_{i+1}$  and  $l$  –  $th$  path from the set connecting the other observer with reference observer (i.e.  $o_1, o_{j+1}$ ) will be the fastest routes, see Fig. 3 for details. Unfortunately those probabilities are non-trivial to evaluate so we assume all paths to be equally probable (which is true only when there are no intersection between them). With this assumption, we obtain

$$\Sigma_{i,j} = \begin{cases} \frac{\sigma^2}{n \cdot m} \sum_{k=1}^n \sum_{l=1}^m |R_k(o_1, o_{i+1}) \cap R_l(o_1, o_{j+1})| & i \neq j \\ \min_{\sigma^2}(\heartsuit_{i+1}) & i = j \end{cases} \quad (19)$$

#### 4.2. Equiprobable links (EPL)

Alternatively one can use a product of Jaccard Index [13] (Fig. 3) and geometric mean of variances of sets of paths, namely:

$$\Sigma_{i,j} = \frac{|\{e_i\} \cap \{e_j\}|}{|\{e_i\} \cup \{e_j\}|} \cdot \sqrt{\frac{\min(\heartsuit_{i+1})}{\sigma^2} \cdot \frac{\min(\heartsuit_{j+1})}{\sigma^2}} \quad (20)$$

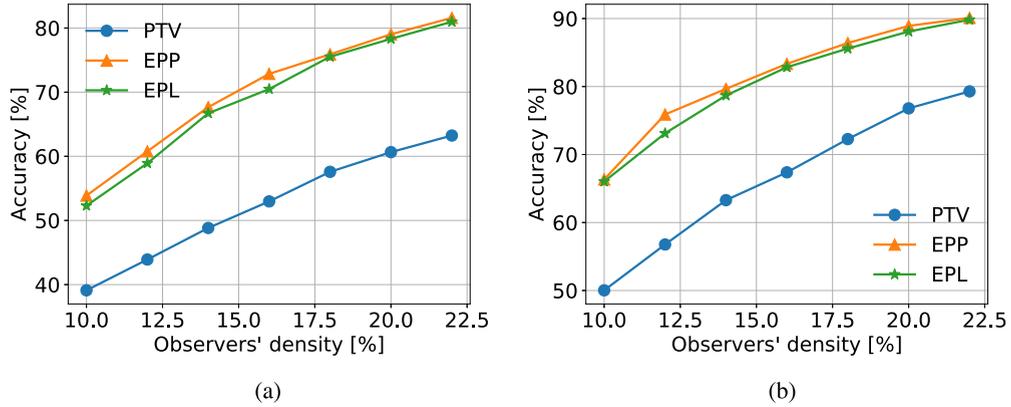
where  $\{e_i\}$  is a set of edges of all shortest paths connecting node  $i$  with the reference observer. Intuitively one could say that using *intersection over union* of sets of edges represents treating each edge as equally probable (instead of treating each *path* equally probable like in previous approach). The choice of geometric mean is arbitrary and arithmetic mean produces similar results.

## 5. Results

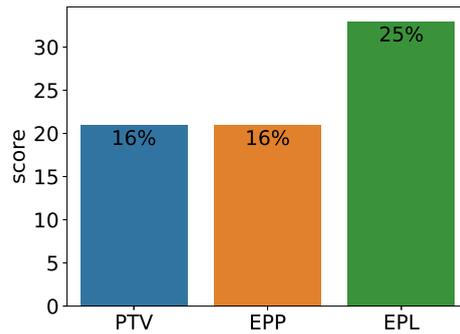
Presented method has been tested on synthetic graphs and real network. We used two synthetic graphs: Barabási–Albert (BA) and Erdős–Rényi (ER) with size  $N = 100$  and average node degree  $\langle k \rangle = 6$  in each case. Since estimator in both methods can produce two or more nodes with the same maximum score the success of algorithm is registered when the actual source is within the list of those candidates with maximum likelihood. Results are shown at Fig. 4. One can clearly see a vast improvement in accuracy of the method using our proposed adjustments with the difference between EPP and EPL being barely noticeable yet still consistently in favour of approach EPP.

We have also conducted tests on a real network: ego-facebook with  $N = 4039$  [14]. After 133 simulations at observers' density  $d = 10\%$  EPL successfully located the source 33 (25%) times while PTV and EPP both scored 21 (16%) (Fig. 5). In this case the EPP modification performed worse than EPL, on the contrary to the case of synthetic networks. The reason for that maybe high clustering coefficient of the real network while ER and BA models are known to not reproduce that characteristic. A high clustering coefficient naturally leads to more correlations between paths and those make our approximation of equally probable paths in approach EPP not sustainable.

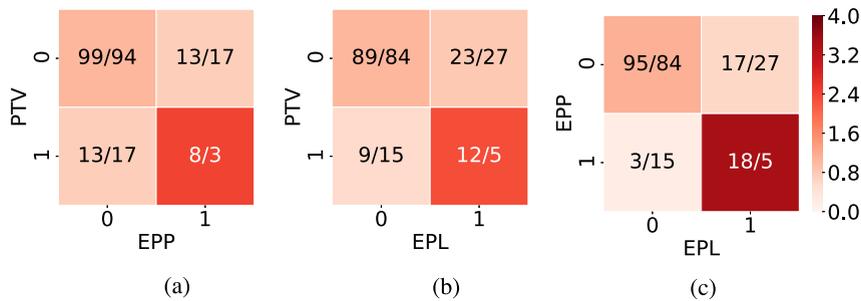
Contingency matrices for a PTV, EPP, EPL methods are presented at Fig. 6. We see that the probability of agreement of any pair of those three methods in successful source location is much higher than in the case they were completely independent.



**Fig. 4.** Improved accuracy of source detection measured as a function of observers' density for original PTV method and our modifications. Simulations were conducted on (a) Barabási-Albert (BA) and (b) Erdős-Rényi (ER) graph. Both with mean node degree  $\langle k \rangle = 6$  and network size  $N = 100$ . We have conducted 100 simulations for each point on those plots.



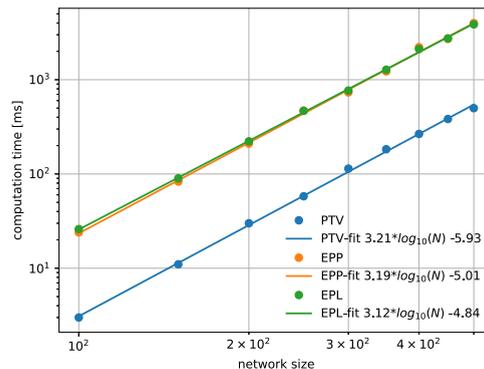
**Fig. 5.** Improved accuracy on real life network. The simulations were conducted on the ego-facebook graph with original PTV method and proposed modifications – EPP and EPL. While PTV and EPP have successfully detected the epidemic source 16% (21/133) of the time the EPL approach has scored 25% (33/133).



**Fig. 6.** Contingency matrices. 0 means the algorithm failed to locate the source, 1 means it succeeded. The first number of the label is the count of appropriate events while the second is the expected based on methods' performance and that they are independent. Colours are applied according to the fractions of the aforementioned numbers.

### 6. Computational complexity

Pinto et al. report that the computational time in their method scales with network size  $N$  as  $N^\alpha$  where for arbitrary graphs  $\alpha = 3$ . In our implementation of PTV we have a reasonably close result of  $\alpha = 3.21$ . Tests were conducted on a Barabási-Albert graph with observers' density  $d = 0.1$ . Our own modifications seem to slightly improve scalability, i.e.  $\alpha_{EPP} = 3.19, \alpha_{EPL} = 3.12$ . However, both EPP and EPL bring significant initial costs resulting in overall computation time to be significantly higher than original PTV's. See Fig. 7 for details. The nature of our modifications also makes theoretical predictions of scalability much harder to acquire. Although in BA example they seem to follow  $N^\alpha$  fit, they heavily depend



**Fig. 7.** Computation times as a function of network size for original PTV and our modifications. Plot is in logarithmic scale and appropriate linear fitting has been done. Simulations were conducted on Barabási–Albert graph with observers' density  $d = 0.1$ .

on graph's density, number of shortest paths and correlations between them. As such it might be advisable to do some pre-computing when applying EPP/EPL on real networks. If observers are known beforehand and do not change one would have to build covariance matrix and deterministic delay only once. The only element changing would be the observed delay vector and all computation is reduced to a vector–matrix–vector multiplication.

## 7. Conclusions & discussion

We have shown that the approach of breadth-first search (*BFS*) trees presented by Pinto et al. overestimates propagation time of a signal. We presented an analysis of why that is, namely there can be more than one shortest paths and also other paths can also happen to be faster, and those effects are non-negligible. We provide alternatives to *BFS* that take multiple paths into account and while requiring more computation time, they significantly improve accuracy of the source detection. Improvement in accuracy is mostly prominent for artificial networks (BA, ER), however, when tested on a real network (ego-facebook) there was a visible increase in accuracy (1.6 times higher) with approach EPL while EPP was no worse than original PTV method. While results of presented methods show a lot of promise there are some obvious paths one can still undertake to improve the accuracy. Firstly, we would love to have an analytical expression for the minimum distribution of an arbitrary amount of Gaussian distributions and not only two. Secondly, one could expect that should the probabilities in EPP variant be known (instead assuming they are all equal) the difference in accuracy between EPP and EPL should increase in the favour of the former. Thirdly, for arbitrary graphs the proposed variants do not scale as trivially with network size as original PTV ( $N^\alpha$ ,  $\alpha \in [3, 4]$ ) for they heavily depend on number of edges, shortest paths and intersections among them thus an exact expression for that scalability is unknown. For both methods, appropriate pre-computing can cut on time needed to locate source after observer reports are obtained.

## Acknowledgments

The work was partially supported as RENOIR Project by the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska–Curie grant agreement No 691152 and by Ministry of Science and Higher Education (Poland), grant Nos. W34/H2020/2016, 329025/PnH /2016 and National Science Centre, Poland Grant No. 2015/19/B/ST6/02612. J.A.H. was partially supported by the Russian Scientific Foundation, Agreement #17-71-30029 with co-financing of Bank Saint Petersburg.

## References

- [1] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. A* 115 (772) (1927) 700–721.
- [2] G. Strauss, A. Shell, R. Yu, B. Acohido, SEC, FBI probe fake tweet that rocked stocks, in: *USA Today*, 2013.
- [3] Romualdo Pastor-Satorras, Alessandro Vespignani, Epidemic spreading in scale-free networks, *Phys. Rev. Lett.* 86 (2001) 3200–3203.
- [4] Mengfeng Sun, Hai-Feng Zhang, Huiyan Kang, Guanghu Zhu, Xinchu Fu, Epidemic spreading on adaptively weighted scale-free networks, *J. Math. Biol.* 74 (09) (2016).
- [5] Christopher Moore, M.E.J. Newman, Epidemics and percolation in small-world networks, *Phys. Rev. E* 61 (2000) 5678–5682.
- [6] Devavrat Shah, Tauhid R. Zaman, Rumors in a network: Who's the culprit? *IEEE Trans. Inf. Theory* 57 (09) (2011) 5163–5181.
- [7] Pedro C. Pinto, Patrick Thiran, Martin Vetterli, Locating the source of diffusion in large-scale networks, *Phys. Rev. Lett.* 109 (2012) 068702.
- [8] Zhesi Shen, Cao Shinan, Wen-Xu Wang, Zengru Di, H. Stanley, Locating the source of diffusion in complex networks by time-reversal backward spreading, *Phys. Rev. E* 93 (03) (2016) 032301.
- [9] Qiangjuan Huang, Chengli Zhao, Xue Zhang, Dongyun Yi, Locating the source of spreading in temporal networks, *Physica A* 468 (2017) 434–444.
- [10] Xizhe Zhang, Yubo Zhang, Tianyang Lv, Ying Yin, Identification of efficient observers for locating spreading source in complex networks, *Physica A* 442 (2016) 100–109.

- [11] Robert Paluch, Xiaoyan Lu, Krzysztof Suchecki, Bolesław K. Szymański, Janusz A. Hołyst, Fast and accurate detection of spread source in large complex networks, *Sci. Rep.* 8 (1) (2018) 2508.
- [12] Saralees Nadarajah, Samuel Kotz, Exact distribution of the max/min of two gaussian random variables, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 16 (03) (2008) 210–212.
- [13] P. Jaccard, Le coefficient generique et le coefficient de communaute dans la flore marocaine, *Mém. Soc. Vaud. Sci. Nat.* 2 (1926) 385–403.
- [14] Jim Mcauley, Jure Leskovec, Learning to discover social circles in ego networks, in: *NIPS*, vol. 1, 2012, pp. 539–547, 01.