Detection of spread source in complex networks

Robert Paluch[1], Łukasz Gajewski[1], Xiaoyan Lu[2], Krzysztof Suchecki[1], Bolesław K. Szymański[2,3], and Janusz A. Hołyst[1,4]

[1]Center of Excellence for Complex Systems Research, Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00662 Warsaw, Poland
[2]Social Cognitive Networks Academic Research Center, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180-3590 USA
[3]The ENGINE Centre, Wroclaw University of Science and Technology, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
[4]ITMO University, 49 Kronverkskiy av., 197101 Saint Petersburg, Russia

Spread over complex networks is a ubiquitous process with increasingly wide applications. Locating spread sources is often important, e.g. finding the patient one in epidemics, or source of rumor spreading in social network. Here we explore the problem of complexity of currently known methods as well as we investigate the validity of the assumption that information spreads only via the shortest paths.
Pinto, Thiran and Vetterli introduced an algorithm (PTVA) to solve the problem of source detection in which a limited set of nodes act as observers and report times at which the spread reached them. PTVA uses all observers to find a solution and assumes the information travels via a single, shortest path, which by assumption is the fastest way. Here we propose a new approach in which observers with low quality information (i.e. with large spread encounter times) are ignored and potential sources are selected based on the likelihood gradient from high quality observers. The original complexity of PTVA is $O(N^{\alpha})$, where $\alpha \in (3, 4)$ depends on the network topology and number of observers (N denotes the number of nodes in the network). Our Gradient Maximum Likelihood Algorithm (GMLA) reduces this complexity to $O(N^2 \log(N))$ without reduction of the detection accuracy.
We also show that assumption that information spreads only via the shortest paths 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.

Acknowledgements

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. X.L. and B.K.S. were partially supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 (the ARL Network Science CTA) and by the Army Research Office grant W911NF-16-1-0524. B.K.S was partially supported by the National Science Centre, Poland, project no. 2016/21/B/ST6/01463. This research was also supported in part by PLGrid Infrastructure.

References

[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, 2508

[2] Gajewski, Ł. G., Suchecki, K., Hołyst, J. A. (2019) Multiple propagation paths enhance locating the source of diffusion in complex networks. *Physica A*, 519, 34