Finding the source of local information cascades in complex networks

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

The issue of finding a source of epidemic or information spreading process in complex network representing society is obviously of a great interest. In most realistic situations we know state of only few select *observers*, not all individuals. The problem has been posed and solved for Susceptible-Infected (SI) type model by Pinto et. al. [1] by using maximum likelihood and treelike approximation. More realistic situations will have connections of different weights and Susceptible-Infected-Resistant type model, where the infection is not guaranteed given time.
We have considered SIR model on weighted, directed network, where weights represent inverse probability of infection (i.e. mean time to transmit). We assume we only know exact network topology (including connection weights) and times at which certain nodes (observers) became infected. We take approach similar to [1], in that we consider each node in the network as potential source, assume treelike approximation of spreading paths and calculate probability distribution of times of infection of observers if given node was the source. In SIR model, information cascades are limited in size and some observers may remain uninfected. Limiting the information to the infected observers we can evaluate likelihood of each node being source as the probability of obtaining exact observed times from distribution associated with that node. This however, would ignore information about where the infection did *not* arrive at all. We include this, by calculating probability $P_r(s, \vec{r})$, that given source s, nodes were reached or not ($\vec{r}$) as observed. This probability is calculated using same treelike approximation for spreading paths, probabilities to infect along each connection ij and probability to become resistant (parameter).
While the likelihood derived from times locates the source in global epidemic cascades, for local cascades the reach likelihood is mostly responsible for success (see Figure).

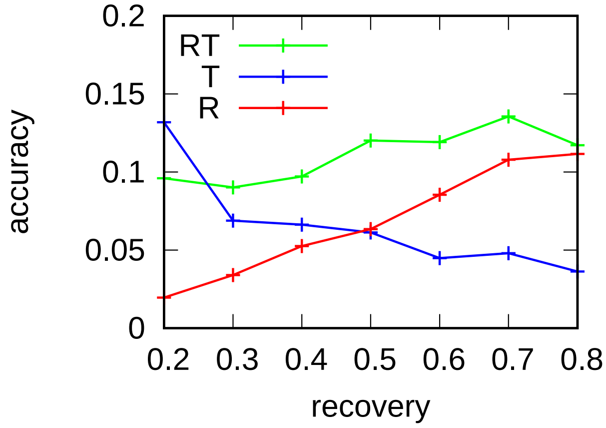

Figure 1. Accuracy (fraction of realizations that found source correctly) of finding the source of SIR process in random network, using only times (T), reach (R) or both (RT), for varying recovery probability. Data for network of $N = 100$ nodes of mean degree $k = 2$, infection rate $I = 0.3$ with $K = 20$ observers and 1000 realizations.

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

References

[1] Pinto, P. C., Thiran, P., & Vetterli, M. (2012) Locating the source of diffusion in large-scale networks. *Physical Review Letters*, 109, 068702