Temporal network compression via network hashing.

Rémi Vaudaine[1], Pierre Borgnat[2], Paulo Goncalves[1], Rémi Gribonval[1], and Márton Karsai[3,4]

[1] Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP, F-69342, Lyon Cedex 07, France
[2] CNRS, Univ de Lyon, ENS de Lyon, Laboratoire de Physique, F-69342 Lyon, France
[3] Department of Network and Data Science, Central European University, 1100 Wien, Austria
[4] Rényi Institute of Mathematics, 1053 Budapest, Hungary

Abstract

Temporal networks are a good representation of an interacting system evolving over time. Dynamical processes on these networks, such as epidemic spreading, burstiness of human interaction or percolation have been particularly studied. But, usually, heavy computations are required to simulate dynamical processes on networks [4] [2] and to extract relevant information. The main issue is that it is computationally very costly since simulations require many draws to evaluate a quantity of interest. A solution to avoid this cost is to use the structure of the network instead of simulations to extract information.

Here the quantities of interest are directly linked to reachability: whether two nodes are connected or not [1] [3]. In temporal networks, paths between pairs of nodes must respect time. Indeed, interactions have to be in the right time-order for an effect to propagate from nodes to nodes. Thus, a time-respecting path is a sequence of adjacent events where two events are adjacent if they share at least one node and the first one occur before the second one. Then, the *out-component of a node* can be defined as the set of nodes reachable by any time-respecting path in the temporal network starting from that specific node. Moreover, the *size of the out-component*, *i.e.* the number of nodes reachable from a source node, is the maximum number of nodes that can be involved in a dynamical process starting from this source. For example, in epidemics spreading, it would be the maximum number of infected people if a virus propagates from a specific node. Finally, the distribution of the size of the out-components is the probability distribution of the out-component's size starting from a random node of the network drawn uniformly among all possible nodes.

To compute the out-components efficiently, we propose a *streaming matrix algorithm* that only requires one scan through the series of events. Then, we also propose a *general purpose compression scheme via hashing* to further improve our algorithm.

References

1. Badie-Modiri, A., Karsai, M., Kivelä, M.: Efficient limited-time reachability estimation in temporal networks. Phys. Rev. E **101**, 052303 (May 2020). https://doi.org/10.1103/PhysRevE.101.052303, https://link.aps.org/doi/10.1103/PhysRevE.101.052303 1
2. Bochenina, K., Kesarev, S., Boukhanovsky, A.: Scalable parallel simulation of dynamical processes on large stochastic kronecker graphs. Future Generation Computer Systems **78**, 502–515 (2018). https://doi.org/https://doi.org/10.1016/j.future.2017.07.021, https://www.sciencedirect.com/science/article/pii/S0167739X17301504 1
3. Holme, P., Saramäki, J.: Temporal networks. Physics Reports **519**(3), 97–125 (oct 2012). https://doi.org/10.1016/j.physrep.2012.03.001, https://doi.org/10.1016%2Fj.physrep.2012.03.001 1
4. Rocha, L.E.C., Liljeros, F., Holme, P.: Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. PLOS Computational Biology **7**(3), 1–9 (03 2011). https://doi.org/10.1371/journal.pcbi.1001109, https://doi.org/10.1371/journal.pcbi.1001109 1