Influence maximization in Boolean networks

Thomas Parmer and Filippo Radicchi

Center for Complex Networks and Systems Research,

Luddy School of Informatics, Computing, and Engineering,

Indiana University, Bloomington, Indiana 47408, USA

Luis M. Rocha

Consortium for Social and Biomedical Complexity,

Systems Science and Industrial Engineering Department,

Thomas J. Watson College of Engineering and Applied Science,

Binghamton University (State University of New York), Binghamton, New York 13902, USA and

Instituto Gulbenkian de Ciência, Oeiras 2780-156, Portugal

The optimization problem aiming at the identification of minimal sets of nodes able to drive the dynamics of Boolean networks toward desired long-term behaviors is central for some applications, as for example the detection of key therapeutic targets to control pathways in models of biological signaling and regulatory networks. Unfortunately, the complexity of the optimization problem is exponential, making it exactly solvable on very small systems only. Some scalable approaches exist but they rely on linear approximations; other approaches estimate nonlinear effects but they are generally not scalable. In this talk, we introduce an alternative method inspired by those used in the solution of the well-studied problem of influence maximization for spreading processes in social networks. The computational time of the proposed method scales cubically with the network size. This is achieved thanks to some strong approximations, as for example neglecting dynamical correlations among Boolean variables. However, the method has the desirable feature of fully accounting for the nonlinear nature of Boolean dynamics. We validate the method on small gene regulatory networks whose dynamical landscapes are known by means of brute-force analysis (Figure 1). We then systematically apply it to a large collection of gene regulatory networks revealing that for about 65% of the analyzed networks, the minimal driver sets contain less than 20% of their nodes.
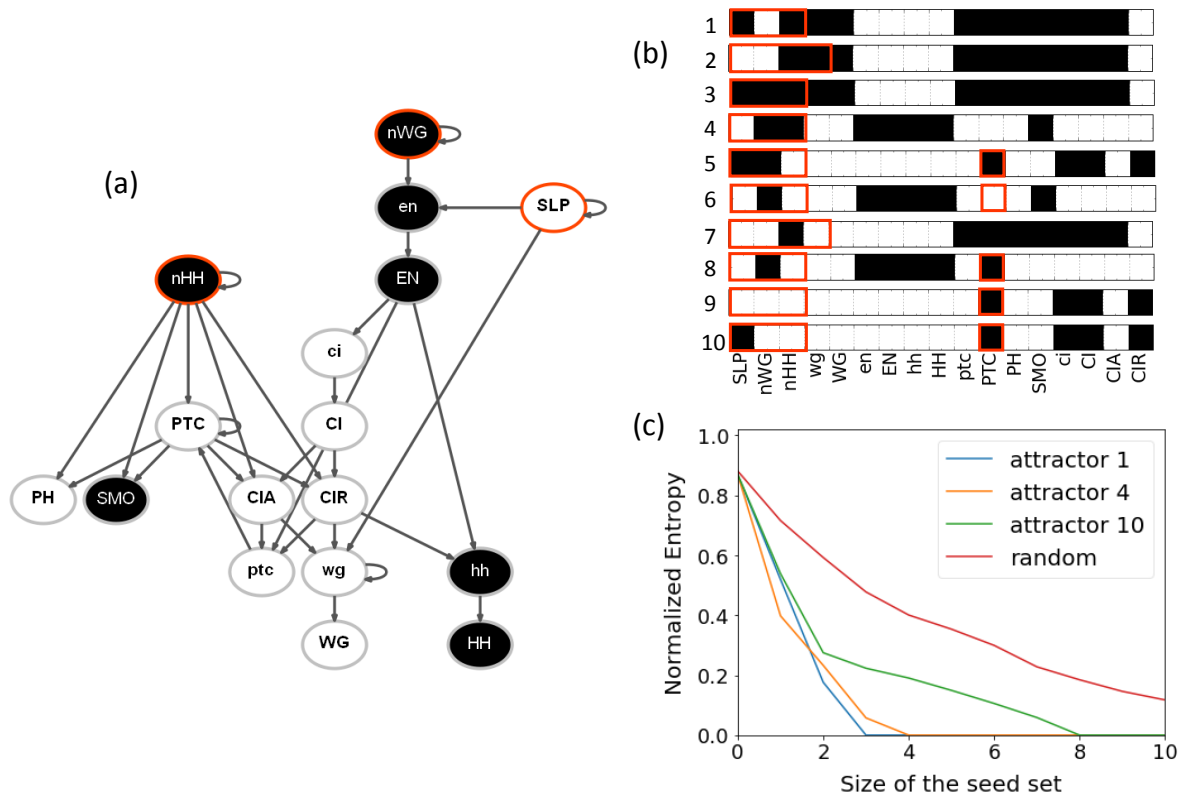
Figure 1: Driving Boolean networks to the desired attractor. (a) We consider the *Drosophila melanogaster* segment polarity network. In the visualization, a directed connection from node i to node j indicates that the state of node i is one argument of the Boolean function F_j that regulates the dynamical evolution of node j. In the visualization, we represent one of the attractors of the network. Active nodes are represented in black; inactive nodes are depicted in white. The attractor may be reached by controlling the state of the three nodes highlighted in red. We identified the seed set using the proposed algorithm; the seed set is $\mathcal{X} = \{(\mathrm{SLP}, \hat{\sigma}_{\mathrm{SLP}} = 0), (\mathrm{nWG}, \hat{\sigma}_{\mathrm{nWG}} = 1), (\mathrm{nHH}, \hat{\sigma}_{\mathrm{nHH}} = 1)\}$. (b) The 10 attractors of the network, and their corresponding minimal driving sets as identified by the proposed algorithm. Attractor 4 is the same as in panel a. Our predictions recover exactly the minimal driver sets for 7 of the 10 fixed points. We overstimate the size of the driver sets required to reach attractors 5, 9 and 10 by one node only (the PTC node). (c) Residual entropy as a function of size of the seed set for three selected seed sets leading to specific attractors. The labels of the attractors are the same as in panel b. The unconstrained greedy selection process finds attractor 4. As a term of comparison, we display also the curve corresponding to seed sets composed of randomly selected nodes' indices/states.