Unraveling the Role of Node Metadata in Network Robustness: the Feature-Based Percolation Model

Oriol Artime[1], and Manlio De Domenico[2]
[1]Bruno Kessler Foundation, Via Sommarive 18, 38123 Povo, Italy
[2]Department of Physics and Astronomy, University of Padua, Via F. Marzolo 8, 35131 Padua, Italy

Percolation is an emblematic model used to understand the robustness of interconnected systems. Despite being a model broadly studied in statistical physics and mathematics, from a theoretical perspective it is usually investigated in relatively simple scenarios, such as the removal of the system's units in random order —simulating unpredictable site failures— or sequentially ordered simulating targeted attacks by specific topological descriptors, the simplest one being the number of node connections. However, in the vast majority of empirical applications, it is required to dismantle the network following more sophisticated protocols than the aforementioned ones, such as based on more convoluted topological properties or even non-topological node metadata obtained from the application domain.

In this work we propose a novel mathematical framework to fill this gap: a network is enriched with features and its nodes are then removed according to their importance in the feature space. Percolation analysis is performed, theoretically and numerically, as a function of the feature distribution, finding an excellent match between the analytical results and the simulations. Several degree-feature relations of diverse nature are explored to show the applicability of the theory. We start from ad hoc degree-feature distributions that capture the main characteristics of correlations observed in empirical systems, moving to features that arise naturally in the process of network creation and ending with the case in which features are coupled to dynamical processes running on top of the network, such as epidemics or biochemical dynamics, among others. Both synthetic and real-world networks of different nature are considered in the analysis. Moreover, we show the potential of our model by employing state-of-the-art Bayesian probability techniques that are able to give the most plausible closed-form expression for the degree-feature distribution when it cannot be computed analytically. By feeding these most plausible expressions into the equations of our model, we can study feature-based percolation in systems for which it is only known the feature and the degree of the individual nodes, instead of the entire degree-feature joint probability distribution (see Fig. 1). This considerably broadens the applicability of the theory and bridges our theory, grounded on statistical physics, with Bayesian machine learning techniques suitable for knowledge discovery.

The article has been recently published in *Percolation on feature-enriched interconnected systems*, Nature Communications, **12**(1), 1-12 (2021).

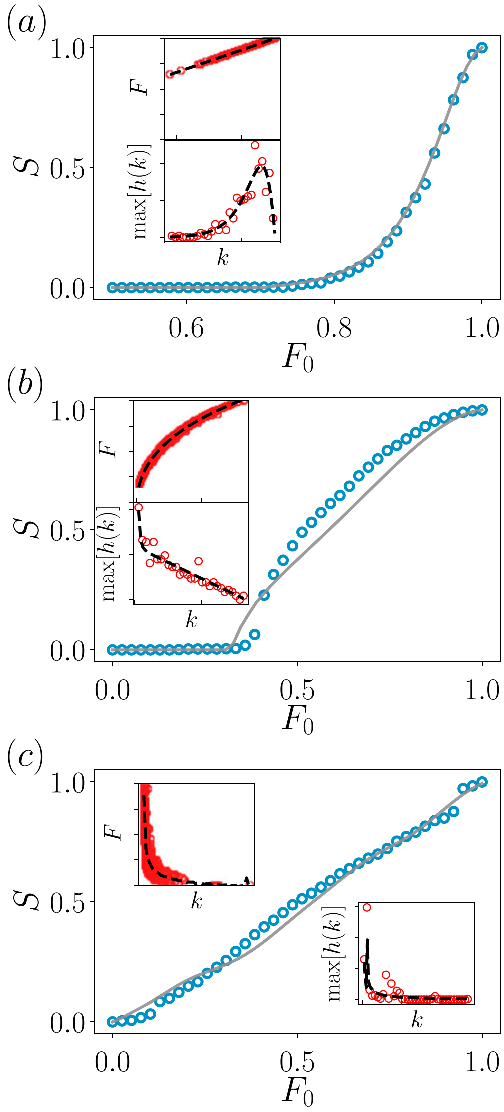

Fig. 1. Percolation curves for (a) mutualistic dynamics in symbiotic ecosystems, (b) population dynamics and (c) mass-action kinetics in biochemistry (solid lines are the theory, points are simulations). F_0 is a parameter of the feature distribution, S is the size of the giant component (order parameter of the phase transition). The insets are different variables fitted using machine-learning techniques that are necessary to compute S. The dynamical equations are not given due to space constrains. All dynamics are simulated in empirical networks.