Percolation in the presence of eigenvector localization

Leo Torres — `leo@leotrs.com`

Max Planck Institute for Mathematics in the Sciences

The concept of percolation threshold has multiple applications in graph theory and network science. One way of studying the percolation threshold is to use a "susceptibility function", which is any function designed to have a tall peak at the percolation threshold. One example is $S(p) \coloneqq \frac{\langle S_p^2 \rangle - \langle S_p \rangle^2}{\langle S_p \rangle}$, where S_p is the size of the largest connected component in the network after a random fraction p of edges has been removed, and the angle brackets denote average over random realizations. Usually, the function $S(p)$ has a single peak that coincides with the percolation threshold of the network.

Another way of estimating the percolation threshold of a network is to use the reciprocal of the leading eigenvalue λ_1 of the *non-backtracking* (NB) matrix (Karrer et al., PRL, 113(208702), 2014). The NB matrix of a graph G is the zero-one matrix that indicates the possible steps taken by a random walker that does not backtrack. This spectral estimate is known to be a lower bound to the true percolation threshold.

In (Pastor-Satorras & Castellano, Sci. Rep. 10.1, 2020), it was pointed out that sometimes the susceptibility function S contains not one, but two peaks, and in this case, the estimate of $1/\lambda_1$ only estimates the first peak and fails to estimate the second peak. It was also pointed out that this seems to happen when the eigenvector corresponding to λ_1 is *localized*, i.e., when there exists a small subgraph H whose nodes receive most of the mass of the eigenvector, while nodes outside of H receive relatively negligible mass.

The open question remains of how to estimate the second peak of $S(p)$ in the presence of eigenvector localization. In this project, I show how to do this. First, I propose the hypothesis that the subgraph H where the eigenvector is localized is percolating at a different rate than the rest of the graph. To test this hypothesis, I propose a greedy algorithm to automatically find H, which is shown to be effective on different real-world data sets (Fig.1a). Second, I show that the reciprocal of the leading eigenvalue of the complement of H is a good estimate for the second peak of $S(p)$, while the reciprocal of the leading eigenvalue of H is a good estimate of the first peak, as usual (Fig.1b). This supports the hypothesis. Third, I show a growth model that is able to generate susceptibility curves with two peaks after being executed on an existing network (Fig.1c). This is derived from first principles and shown to be a mixture between preferential attachment and copying strategies.

Finally, both the greedy algorithm and the growth process are based on the observation in (Torres et al., SIMODS, 3(2), 2021) that the higher the *X-Degree centrality* of a node, the higher its influence on λ_1. The X-Degree centrality is derived from a perturbation analysis of the NB matrix and is a quadratic aggregation of a node's neighbors' degrees. Of note, X-Degree is purely a local and structural centrality measure, and it is observed here to effectively manipulate a global, spectral phenomenon such as eigenvector localization.

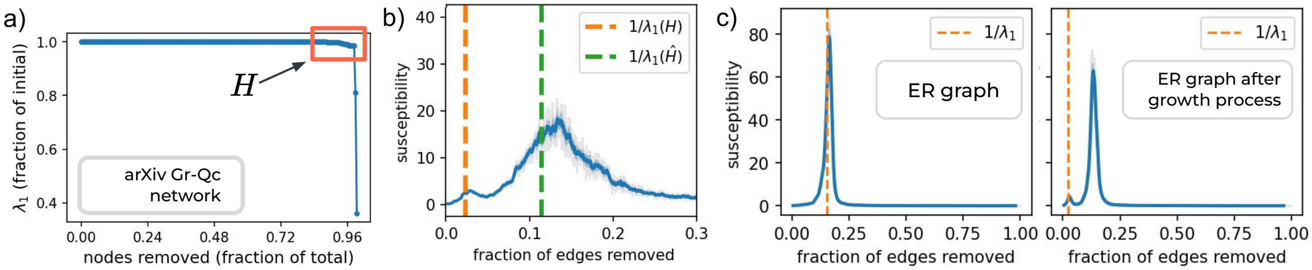

Figure 1. a) Removing nodes based on X-Degree reveals a sharp drop in λ_1 in a collaboration network. We choose H as the smallest subgraph such that removing any more nodes causes a drastic change in λ_1. b) The eigenvalues of H and its complement $\hat{H}$, when taken separately, serve to estimate the two peaks in the susceptibility function in the original network. c) The susceptibility curve of a random ER graph before (left) and after (right) the growth process based on X-Degree has been applied.