

## A network-specific approach to percolation in complex networks with bidirectional links

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2012 EPL 98 16007

(<http://iopscience.iop.org/0295-5075/98/1/16007>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 128.138.249.14

The article was downloaded on 07/05/2012 at 05:47

Please note that [terms and conditions apply](#).

# A network-specific approach to percolation in complex networks with bidirectional links

DANE TAYLOR<sup>(a)</sup> and JUAN GABRIEL RESTREPO

*Department of Applied Mathematics, University of Colorado - Boulder, CO 80309, USA*

received 11 January 2012; accepted in final form 8 March 2012

published online 12 April 2012

PACS 64.60.ah – Percolation

PACS 89.75.-k – Complex systems

PACS 05.70.Jk – Critical point phenomena

**Abstract** – Methods for determining the percolation threshold usually study the behavior of network ensembles and are often restricted to a particular type of probabilistic node/link removal strategy. We propose a network-specific method to determine the connectivity of nodes below the percolation threshold and offer an estimate to the percolation threshold in networks with bidirectional links. Our analysis does not require the assumption that a network belongs to a specific ensemble and can at the same time easily handle arbitrary removal strategies (previously an open problem for undirected networks). In validating our analysis, we find that it predicts the effects of many known complex structures (*e.g.*, degree correlations) and may be used to study both probabilistic and deterministic attacks.

Copyright © EPLA, 2012

**Introduction.** – The study of percolation in complex networks has broad applications including epidemic spreading [1], propagation of excitation in neural networks [2], and robustness of networks to random failure [3] or strategic attack [4,5]. A central problem is estimating the *percolation threshold*, the critical fraction of nodes or links of an initially connected network that must be removed to disintegrate it into small disconnected fragments. Knowledge of how a network fragments can improve strategies for designing attack [4,5] and immunization techniques [6] or increasing network robustness [7,8].

Several studies have proposed techniques to estimate the percolation threshold of a network for various situations [4,9–13]. These studies typically use *ensemble approaches*, where one studies the typical behavior of a set of networks satisfying some set of properties. Common ensembles include: i) networks with fixed degree sequence generated from, for example, the configuration model [14]; ii) networks with an expected degree sequence generated with, for example, the Chung-Lu model [15]; and iii) *Markovian* networks with correlations between nearest neighbors, where correlations may be captured by the probability  $P(\mathbf{d}'|\mathbf{d})$  that a node with degree  $\mathbf{d}$  is connected to a node of degree  $\mathbf{d}'$ . Here  $\mathbf{d} = (d^{in}, d^{out})$  denotes the number of incoming ( $d^{in}$ ) and outgoing ( $d^{out}$ ) links at a

given node. While significant progress has been made in the study of such ensembles [4,9–12], we note two important limitations of the ensemble approach. i) Given a single network, it is not clear what ensemble should be selected to capture the properties of the network. Typically, networks found in applications contain various structural properties (*e.g.*, correlations [12], clustering [16], and community structure [17]) that are not always accounted for in the ensembles. A related problem is that it has been recently observed [18] that, given an ensemble of networks, some network properties can vary significantly within the ensemble. Thus, it is not clear that ensemble approaches give the best description of a single network found in practice. ii) Some ensemble theories are impractical when applied to individual real networks. For example, theories to estimate the percolation threshold in Markovian ensembles [11] require the estimation of a potentially dense and very large,  $(d_{max}^{in} d_{max}^{out}) \times (d_{max}^{in} d_{max}^{out})$  matrix  $P(\mathbf{d}|\mathbf{d}')$ , which is difficult to obtain from a single network.

In this letter, we develop a network-specific theory for percolation in networks with bidirectional links. In contrast to previous approaches to network-specific estimates of the percolation threshold which were restricted to purely directed networks [13], our work has a broader range of applicability as it can be applied to undirected networks and directed networks with bidirectional links. Our method is based on an analysis of the network's adjacency matrix, which is often known or can be

<sup>(a)</sup>E-mail: dane.taylor@colorado.edu

estimated in important applications (*e.g.*, the power grid [19] and air transportation networks [20]). Besides relaxing the ensemble assumptions of previous research (*e.g.*, that the network is strictly Markovian), one significant advantage of this approach is that it can easily account for arbitrary strategies of node/link removal. Network-specific approaches are therefore well suited for developing network-specific attack/defense strategies, immunization techniques, etc. In addition to estimating the percolation threshold, we predict the expected number of nodes accessible to each node after the network disintegrates. This has various applications such as predicting the outbreak size of an epidemic [1]. We finally show that our method may be used to study the fragmentation of a network subject to either probabilistic or deterministic attack.

**Analysis.** – We formalize weighted percolation (*i.e.*, in which nodes and/or links are retained with arbitrary probabilities) as follows: for a network with  $N$  nodes described by a possibly asymmetric adjacency matrix  $A$  ( $A_{nm} = 1$ , if a link exists from node  $n$  to node  $m$  and  $A_{nm} = 0$ , otherwise), node  $n$  is retained with probability  $q_n$ , and the directed link from node  $n$  to node  $m$  is retained with probability  $p_{nm}$ . Letting  $q = N^{-1} \sum_n q_n$  denote the *average node retention probability*,  $qN$  nodes are expected to remain after a realization of this process (referred to as a *percolation trial*). Unweighted node percolation corresponds to  $q_n = q < 1$  and  $p_{nm} = 1$ , while unweighted link percolation corresponds to  $q_n = 1$  and  $p_{nm} = p < 1$ . For our analysis, it is useful to introduce a matrix  $\hat{A}$  with entries defined by  $q\hat{A}_{nm} = q_m p_{nm} A_{nm}$ , which represents the probability that a link exists from node  $n$  to node  $m$ , given that node  $n$  is retained. Because our analysis depends only on the matrix  $\hat{A}$ , it is applicable to link, node, and mixed (*i.e.*, simultaneous link and node) percolation. However, for the remainder of this letter we consider only node percolation,  $p_{nm} = 1$ .

For a given node-targeting strategy, defined as a set of retention probabilities  $\{q_n\}$ , we are interested in the size  $s$  of the *largest strongly connected component* (LSCC), the largest subset of nodes so that any node in the subset is reachable from any other node in the subset. The percolation threshold  $q^*$  is defined as the value of  $q$  such that  $s \ll N$  for  $q < q^*$  (the *subcritical regime*) and  $s \sim N$  for  $q > q^*$  (the *supercritical regime*). For a particular removal strategy, the subcritical regime may be analyzed by noting that after a percolation trial, only a fraction of the network is reachable from a given node  $n$  following directed links. Following [10] we define, for a given percolation trial, the *out-component* of node  $n$  as the set of nodes that may be reached from node  $n$  via the remaining network (including node  $n$ ) and define  $s_n^{out}$  as the size of the out-component of node  $n$  averaged over many percolation trials. To motivate subsequent analysis, consider first the case when the network is a directed tree. In that case  $s_n^{out}$  satisfies the relation

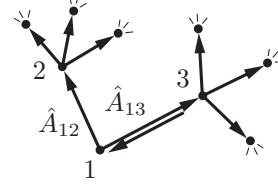


Fig. 1: When computing  $s_1^{out}$ , to compensate for the over-counting of nodes due to the bidirectional link  $1 \leftrightarrow 3$ ,  $\beta_{13}$  reduces the contribution of  $s_3^{out}$  on  $s_1^{out}$ . We approximately have  $\beta_{13} \sim 3/4$  and  $s_1^{out} = 1 + q\hat{A}_{12}s_2^{out} + q\hat{A}_{13}\beta_{13}s_3^{out}$ .

$s_n^{out} = 1 + \sum_m q\hat{A}_{nm}s_m^{out}$ , where the right-hand side counts the nodes reachable from node  $n$  by counting the nodes reachable from its neighbors, and adds 1 to account for node  $n$  itself. The same expression approximately applies to directed networks that are locally tree-like [21] and leads to the results in ref. [13],  $q^* \approx \hat{\lambda}^{-1}$ , where  $\hat{\lambda}$  is the principal, or Perron-Frobenius, eigenvalue of  $\hat{A}$  (*i.e.*,  $\hat{A}\hat{\mathbf{u}} = \hat{\lambda}\hat{\mathbf{u}}$ ). When links are allowed to be bidirectional, however, the expression above overestimates the size of  $s_n^{out}$ , since the terms  $s_m^{out}$  on the right-hand side might include nodes that are reachable by following links back into node  $n$  (see fig. 1).

To correct for this over-counting of nodes, we heuristically modify the contribution of  $s_m^{out}$  on the right-hand side by a factor  $\beta_{nm}$  (to be determined),

$$s_n^{out} = 1 + \sum_m q\hat{A}_{nm}\beta_{nm}s_m^{out}. \quad (1)$$

To determine a self-consistent expression for  $\beta_{nm}$ , we note that from eq. (1) the relative contribution of the out-component of node  $m$  on  $s_n^{out}$  is  $q\hat{A}_{nm}\beta_{nm}s_m^{out}/s_n^{out}$ . Therefore, to reduce the contribution of  $s_m^{out}$  on  $s_n^{out}$  to account for the branch returning to node  $n$  (if present), we let  $\beta_{nm} = 1 - q\hat{A}_{mn}\beta_{mn}s_n^{out}/s_m^{out}$ . Inserting here the corresponding expression for  $\beta_{mn}$  and solving for  $\beta_{nm}$  we obtain

$$\beta_{nm} = (1 - q^2\hat{A}_{nm}\hat{A}_{mn})^{-1} \left( 1 - \frac{q\hat{A}_{mn}s_n^{out}}{s_m^{out}} \right). \quad (2)$$

After substitution of eq. (2) into eq. (1), we find

$$\mathbf{s}^{out} = [I - D(q)]^{-1} \mathbf{y}, \quad (3)$$

where  $\mathbf{s}^{out} = [s_1^{out}, \dots, s_N^{out}]^T$ ,  $I$  is an identity matrix of size  $N$ ,  $\mathbf{y}$  is a vector with entries

$$y_n = \left[ 1 + q^2 \sum_k \hat{A}_{nk}\hat{A}_{kn}(1 - q^2\hat{A}_{nk}\hat{A}_{kn})^{-1} \right]^{-1}, \quad (4)$$

and  $D(q)$  is a matrix with entries

$$D_{nm}(q) = q\hat{A}_{nm}y_n(1 - q^2\hat{A}_{nm}\hat{A}_{mn})^{-1}. \quad (5)$$

Given a removal strategy, eq. (3) can be solved to obtain the expected out-component size for each node. To obtain an estimate for the percolation threshold, note that eq. (3)

requires the invertibility of the matrix  $I - D(q)$ . This matrix is invertible when  $\lambda_{D(q)} < 1$ , where  $\lambda_{D(q)}$  is the principal eigenvalue of  $D(q)$ . As  $\lambda_{D(q)} \rightarrow 1^-$  the out-component sizes diverge as  $s_n^{out} \sim [1 - \lambda_{D(q)}]^{-1} w_n$ , where  $\mathbf{w}$  is the principal eigenvector of  $D(q)$ . A similar argument can be made for the divergence of the in-component sizes. Since the LSCC above the percolation threshold can be thought of as the set of vertices with infinite in- and out-components [11], we predict the percolation threshold as

$$q_D^* = \min_{q \in [0,1]} \{q : \lambda_{D(q)} = 1\}. \quad (6)$$

We note that if there are no bidirectional links,  $\hat{A}_{nm}\hat{A}_{mn} = 0$  and  $D(q) = q\hat{A}$ , and the results of ref. [13] are recovered. While one may solve eq. (6) numerically, it is both practical and insightful to approximate eqs. (2)–(6) for large  $s_n^{out}$  and small  $q$ . Letting  $s_n^{out} \gg 1$  and  $\beta_{nm} \sim 1$  in eq. (1) yields the approximate eigenvalue problem  $s_n^{out} \approx q \sum_m \hat{A}_{nm} s_m^{out}$ . It follows that  $s^{out} \propto \hat{\mathbf{u}}$ . Upon substitution we find  $q \sim \hat{\lambda}^{-1}$  under these conditions, yielding to first order  $\beta_{nm} \approx 1 - \hat{\lambda}^{-1} \hat{A}_{nm} \hat{u}_n / \hat{u}_m$ . Defining

$$C_{nm} = \hat{A}_{nm} \left( 1 - \frac{\hat{A}_{mn} \hat{u}_n}{\hat{\lambda} \hat{u}_m} \right), \quad (7)$$

with principal eigenvalue equation  $C\mathbf{z} = \lambda_C \mathbf{z}$  and using  $\mathbf{y} \approx \mathbf{1} = [1, 1, \dots, 1]^T$ , we obtain the predictions

$$\mathbf{s}^{out} \approx (I - qC)^{-1} \mathbf{1}, \quad (8)$$

$$q_C^* \approx \lambda_C^{-1}. \quad (9)$$

In addition to offering simplified predictions for  $s_n^{out}$  and  $q^*$ , for unweighted percolation (*i.e.*,  $\hat{A} = A$  and  $\hat{\mathbf{u}} = \mathbf{u}$ ) these estimates allow us to bound  $\lambda_C$  using the principal eigenvalue  $\lambda$  of the network adjacency matrix  $A$  (*e.g.*,  $A\mathbf{u} = \lambda\mathbf{u}$ ). Direct application of the Bauer-Fike theorem [22] for the limiting case of an undirected network yields  $|\lambda_C - \lambda| \leq \|\lambda^{-1} U A U^{-1}\|_2 = 1$ , where  $U = \text{diag}[u_1, \dots, u_N]$ . Finally, considering  $\mathbf{1}^T C \mathbf{u}$  and using  $\mathbf{z} \sim \mathbf{u}$  yields  $\lambda_C \approx \lambda - 1$ . One implication of these results is that  $q^* \rightarrow 0$  for large  $\lambda$ , which is consistent with the lack of an unweighted percolation threshold for well-connected networks such as scale-free networks [9]. We note that eqs. (8) and (9) are in best agreement with eqs. (3) and (6) near  $q = q^*$  and when the network is strictly undirected or strictly directed.

**Examples.** – In what follows, we will motivate the need for our theory, recover previous results, explore several applications, and illustrate the robustness of our analysis to complex structures in networks. We will consider both computer-generated and real-world networks.

We first highlight the need for a network-specific method for undirected networks and show that unlike the ensemble approach, a network-specific method captures variability in  $q^*$  across an ensemble. We consider percolation in an uncorrelated, random network formed by

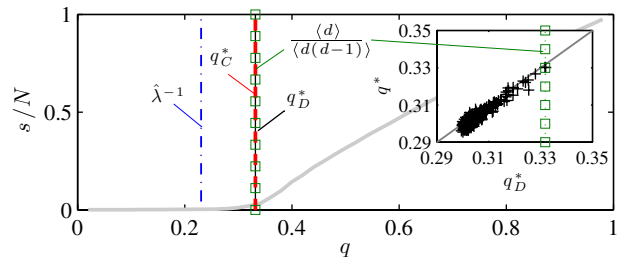


Fig. 2: (Colour on-line) The relative size of the giant component,  $s/N$  (solid grey line), is shown for unweighted percolation on an ER network [23] (see text). Our predictions to  $q^*$  given by eq. (6) and eq. (9) are given by vertical solid black and dashed red lines and are in good agreement with the undirected ensemble result [10] (green squares). The network-specific prediction for directed networks,  $\hat{\lambda}^{-1}$  [13] (blue dot-dashed line), is shown to be inaccurate. The inset shows experimental values for  $q^*$  vs. eq. (6) for networks within an ensemble. Unlike the ensemble approach, variation in  $q^*$  is naturally accounted for by our network-specific approach.

retaining the giant component from an Erdős-Rényi process [23] (henceforth referred to as an ER network) with  $N = 10^4$  nodes and  $3N$  links. In fig. 2 we show the fraction of retained nodes in the LSCC,  $s/N$ , as a function of  $q$  (solid grey curve). Our predictions for  $q^*$  given by eq. (6) (black line) and eq. (9) (red dashed line) and the undirected ensemble theory [10] (green squares) work well, whereas the network-specific theory for directed networks,  $\hat{\lambda}^{-1}$  [13] (blue dot-dashed line), does not, as expected. The inset shows experimental values for  $q^*$  vs. our prediction using eq. (6) for an ensemble of uncorrelated networks obtained by rewiring the ER network while retaining a fixed degree sequence,  $[d_1, d_2, \dots, d_N]$ . The rewiring process is similar to that in refs. [5,12], except an additional step is taken to ensure the resulting network contains all nodes in its LSCC. Note that whereas our network-specific method naturally accounts for variability in  $q^*$  across the ensemble, the ensemble approach (squares) cannot, predicting  $q^* \approx 0.33$  for all members of the ensemble.

Experimental values for  $q^*$  in this inset and remaining figures were found using extrapolation over 6 overlapping intervals of width 0.03 that span  $[0.03, 0.13]$  to find the intercept with the horizontal axis. This yielded a mean and standard deviation (shown when significant). While the actual percolation threshold  $q^*$  is often well defined for model networks in the  $N \rightarrow \infty$  limit (*e.g.*, by using traditional finite-size scaling arguments), one can only estimate it for real-world networks which have a fixed and finite size. We also note that while  $q^*$  can often be extracted by examining the size of the second-largest cluster, this approach was observed to significantly overestimate  $q^*$  for the relatively small networks considered here for which  $N < 10^5$ .

We next illustrate our analysis in networks with bidirectional links and recover previous results for unweighted percolation in uncorrelated networks. Letting  $f$  denote the

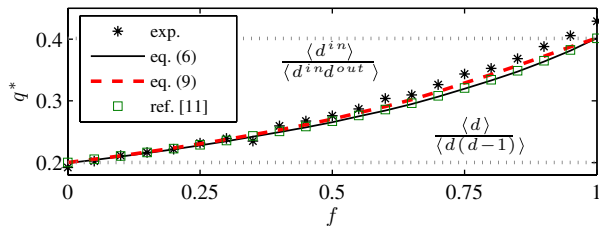


Fig. 3: (Colour on-line) Equation (6) and eq. (9) agree well with ref. [11] and experimental values for predicting the unweighted percolation threshold  $q^*$  of an uncorrelated random network.  $f$  denotes the fraction of links that are directed, which increases from 0 (undirected) to 1 (directed).

fraction of directed links, we begin with an undirected (*i.e.*,  $f = 0$ ) ER network with  $N = 10^4$  nodes and  $5N$  links and iteratively replace randomly chosen undirected links with directed links in a random orientation until the network is strictly directed (*i.e.*,  $f = 1$ ). One can observe in fig. 3 that our predictions by eqs. (6) and (9) agree well with the ensemble result [11] and experimental values. Note that the predictions of eq. (9) agree with those of eq. (6) very well for  $f = 0$  or 1 and that maximal disagreement occurs near  $f \sim 0.7$ . We note that one disadvantage of extrapolating on a fixed range (*e.g.*,  $s/N \in [0.03, 0.13]$ ) while varying  $q^*$  is that a drift occurs due to the combination of finite-size effects and the fact that the slope of an  $s(q)$  curve just above  $q^*$  varies monotonically with  $q^*$ . This is one cause for the disagreement between theory and experiment in fig. 3 for large  $f$ .

We now recover the ensemble predictions for strictly undirected and strictly directed networks shown by horizontal dotted lines in fig. 3. For  $f = 0$ , our result  $\lambda_C \approx \lambda - 1$  and the undirected mean-field (MF) result  $\lambda \approx \langle d^2 \rangle / \langle d \rangle$  lead to  $q^* \approx \langle d \rangle / \langle d(d-1) \rangle$  [9]. For  $f = 1$  we have  $A_{nm}A_{mn} = 0$  and  $C = A$ , which recovers  $q^* \approx \lambda^{-1}$  [13]. Again, the MF result  $\lambda^{-1} \approx \langle d^{in} \rangle / \langle d^{in} d^{out} \rangle$  [13] recovers the result of ref. [10].

Our next examples explore targeted attacks, where we let  $q_n \propto d_n^l$  for  $l \in \{-1, 0, 1\}$ . Here  $d_n = (d_n^{in} + d_n^{out})/2$  denotes the degree of node  $n$  averaged over incoming and outgoing links. Preferentially removing nodes with large degrees (*i.e.*,  $l < 0$ ) can model probabilistic attacks [4] and biased infrastructure failure [3], whereas preferentially removing nodes with small degrees ( $l > 0$ ) may represent a non-obtrusive degradation. We consider real-world networks for which previous methods are either impractical or not justified for the reasons mentioned in the introduction. In fig. 4(a) we show the relative size of the LSCC,  $\phi = s/(qN)$ , for weighted percolation on a directed Word-Association (WA) Network [24]. Vertical dashed lines and  $\times$ 's represent the predictions of eq. (6) and eq. (9). The inset shows this graph with the horizontal axis respectively normalized by the value of  $q^*$  found from eq. (6) for each targeting strategy, where a LSCC of size  $s \sim N$  appears at  $q/q^* \sim 1$  for all curves. In fig. 4(b), we show the prediction by eq. (6) (solid lines) and observed (symbols) values of  $q^*$  as a function of  $l$ , normalized by the value of  $q^*$  at  $l = 0$ , for the WA network [24] and an undirected

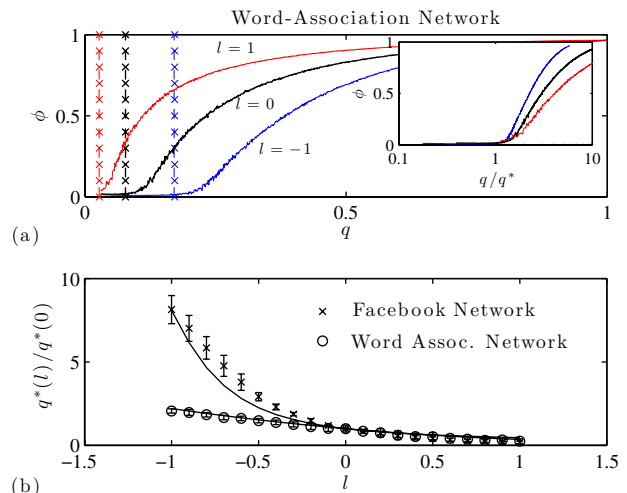


Fig. 4: (Colour on-line) (a) Average values for  $\phi = s/(qN)$  over 32 trials are shown for weighted percolation with  $q_n \propto d_n^l$  on the scale-free WA network [24]. Vertical dashed lines and  $\times$ 's represent the predictions of eq. (6) and eq. (9). The inset shows the same quantities with the  $q$ -axis normalized by eq. (6) for each  $l$  value. (b) The dependency of  $q^*$  on  $l$  is shown for the WA (circles) and FB networks [25] ( $\times$ 's).

network of Facebook (FB) friendships at Caltech [25]. Being a well-connected network with power-law degree distribution, the FB network has a very small unweighted percolation threshold ( $q^*(0) \sim 0.01$ , not shown). However, preferentially removing nodes with large degree can yield non-trivial thresholds with nearly a ten-fold increase ( $q^*(-1) \sim 0.1$ ).

Our next example illustrates eqs. (3) and (8) in a peer-to-peer (P2P) network of file downloads [26] with  $N = 6301$  nodes for weighted site percolation with  $q_n \propto d_n^{-0.5}$ . In this example  $s_n^{out}$  can be interpreted as the expected number of nodes infected by a computer virus released by user  $n$  and preferentially targeting nodes with large degree [1]. In fig. 5 we show a sample of the values of  $s_n^{out}$  predicted by eq. (3) ( $\times$ 's) (which agrees with eq. (8)) *vs.* their experimental values,  $s_n^{out}$  (exp). The prediction given by eq. (3) ( $\times$ 's) is very accurate for the average removal probability  $q = 0.2$  shown in fig. 5(a). For the larger value  $q = 0.4$ , our prediction deviates somewhat from the observed values as shown in fig. 5(b). This is expected because  $s_n^{out}$  is predicted to diverge at  $q^* \sim 0.476$ , but experimental values are bounded by the finite network size  $N$ , so the predicted value must become larger than the observed value as  $q \rightarrow q^*$ .

In our next example we apply our results to network fragmentation under deterministic attack. We consider two important undirected networks, a power grid [19] and an airline transit network [20], and point out that knowledge of how they fragment is essential for taking measures of protection. Simulated attacks were implemented by iteratively removing the node  $n$  corresponding to the largest dynamical importance,  $DI_n \propto u_n v_n$  [5], where  $\mathbf{u}, \mathbf{v}$  are the right and left principal eigenvectors of  $A$ . For these undirected networks,  $DI_n = u_n^2$ . After each node



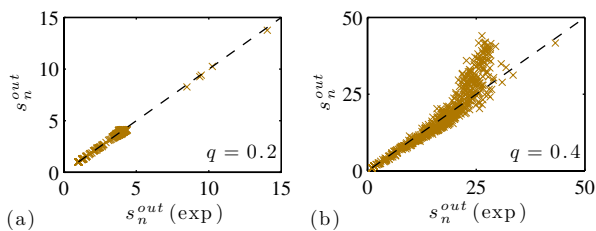


Fig. 5: (Colour on-line) (a) For weighted percolation on a P2P network [26] with retention probability  $q_n \propto d_n^{-0.5}$  and average retention rate  $q = 0.2$ , the values of  $s_n^{out}$  predicted by eq. (3) ( $\times$ 's) agree well with experimental values,  $s_n^{out}(\text{exp})$ , which were averaged over  $2^{16}$  percolation trials. (b) Similar results for  $q = 0.4$ , where some deviations appear (see text).

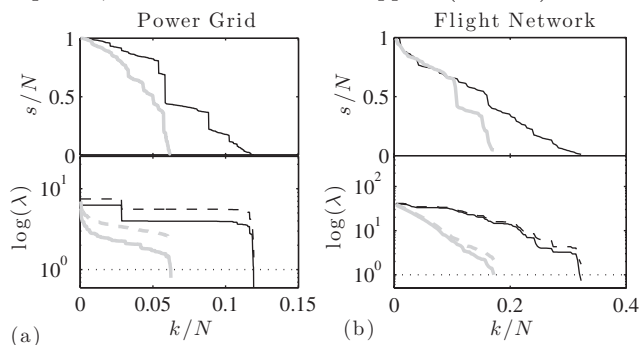


Fig. 6: Malicious attacks were simulated on two significant undirected networks: (a) a power grid for the U.S. [19] and (b) a flight network for US Airlines [20], where the node with largest dynamical importance [5] was iteratively removed (along with nodes disconnected from the LSCC after its removal). The relative size of the LSCC is shown in the top panels (grey lines) as a function of the fraction of removed nodes,  $k/N$ . The relevant eigenvalues for the networks' adjacency matrices, *i.e.*,  $\lambda$  (dashed grey lines) and  $\lambda_C$  (solid grey lines), are shown in the bottom panel as a function of the fraction  $k/N$  of removed nodes. Black lines indicate the same variables under random node removal. Note that in all cases the disappearance of the LSCC corresponds to  $\lambda_C \approx 1$  (horizontal dotted lines).

removal, components fragmented from the LSCC were also removed. In the upper panels of fig. 6, we show the relative size of the LSCC after the removal of  $k$  nodes (grey lines), whereas in the lower panels we show both  $\lambda$  (dashed grey lines) and  $\lambda_C$  (solid grey lines), where  $C$  is given by eq. (7). The horizontal dotted line denotes  $\lambda_C = 1$ . For comparison, black lines indicate the same variables for random node removal. Note that in all cases the disappearance of the LSCC corresponds to  $\lambda_C \approx 1$ . For these undirected networks this corresponds to  $\lambda - 1 \approx 1$ , whereas for directed networks we recover the result of ref. [13],  $\lambda \approx 1$ . Here  $\lambda$  is the principal eigenvalue of the network adjacency matrix  $A$ .

In the remaining examples we demonstrate the robustness of our results to complex structures in networks. Beginning with degree correlations, we first address Markovian correlations between the degrees of adjacent nodes, which can be characterized by the assortativity coefficient  $r \in (-1, 1)$  [12], where  $r > 0$  ( $r < 0$ ) indicates that nodes tend to connect to other nodes with similar

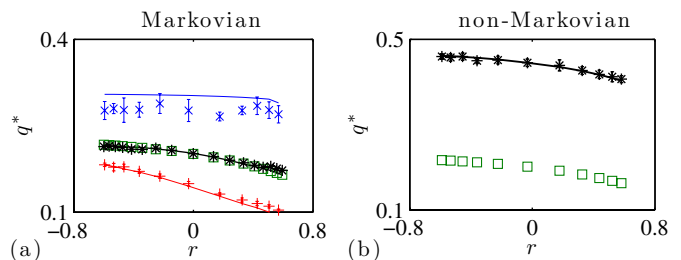


Fig. 7: (Colour on-line) (a) Values of  $q^*$  predicted from eq. (6) (solid lines) and observed experimentally (symbols) for percolation on an ER network having  $N=10^4$  nodes and  $5N$  undirected links rewired with correlations (measured by assortativity coefficient  $r$ ). The three curves correspond to unweighted percolation ( $l=0$ , middle), and weighted percolation ( $l=-1$ , top, and  $l=1$ , bottom). Squares show the undirected Markovian prediction [12]. (b) Equation (6) (solid line) predicts observed values of  $q^*$  (stars) for a non-Markovian network (see text), whereas the Markovian prediction (squares) cannot.

(different) degrees. We again allow the retention probability of a node to depend on its degree,  $q_n \propto d_n^l$ . We note that ref. [4] provides an ensemble approach for a similar removal strategy, but their analysis is restricted only to scale-free networks lacking correlations. In fig. 7(a) we show the effect of degree correlations on  $q^*$  for an ER network with  $N = 10^4$  nodes and  $5N$  undirected links that is rewired to have correlations. Experimental values are shown for  $l = -1$  ( $\times$ 's),  $l = 0$  (stars), and  $l = 1$  (crosses). Solid lines indicate the prediction of eq. (6), which was found to coincide with that of eq. (9), and the squares indicate the undirected Markovian prediction of ref. [12] (applicable only for  $l = 0$ ). Degree correlations were varied while keeping the degree distribution constant following the algorithm in refs. [5,12]. Note that while assortativity promotes robustness for unweighted percolation ( $l = 0$ ) by reducing  $q^*$ , we find that its effect is reduced (amplified) for  $l < 0$  ( $l > 0$ ). For example, the percolation threshold is largely unaffected by degree correlations for  $q_n \propto d_n^{-1}$  for this network (see  $\times$ 's in fig. 7(a)).

Turning to degree correlations of the non-Markovian type in undirected networks, a case which no previous theories can handle, we consider the set of networks used to produce fig. 7(a) but subject them to the following rewiring process: each link  $m \leftrightarrow n$  is replaced by two new links and a new node  $j$ ,  $m \leftrightarrow j \leftrightarrow n$ , resulting in correlations across paths of length two. In fig. 7(b), whose horizontal axis is carried over from fig. 7(a), we show that the network-specific prediction eq. (6) (solid line) agrees with the observed values of  $q^*$  (stars) for unweighted percolation on these non-Markovian networks. For comparison, direct application of the Markovian ensemble method [12] (squares) does not give good results.

We conclude by showing that our results remain accurate for relatively large clustering coefficient,  $c$  [16]. In this experiment links were iteratively added to an ER network by finding paths of length two and completing the triangles. In fig. 8 we compare our predictions with the observed values of  $q^*$  for variable  $c$ .

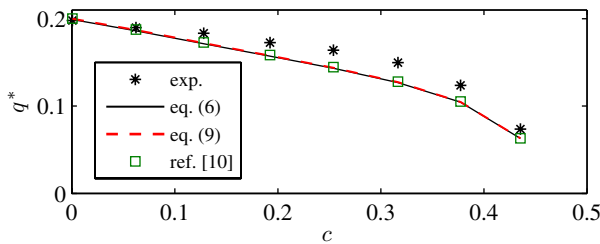


Fig. 8: (Colour on-line) Predictions to  $q^*$  and experimental values for percolation on an ER network with  $N = 10^4$  nodes and  $5N$  links as a function of the clustering coefficient,  $c$  (see text). Note that the lack of degree correlations yields good agreement between our results, eq. (6) and eq. (9), and the ensemble result of ref. [10].

**Conclusions.** – We have presented a *network-specific* approach to weighted percolation in undirected networks and directed networks with bidirectional links (previously open problems). As opposed to most previous theory dealing with network ensembles, our method predicts unique percolation characteristics for each unique network. While ensemble and network-specific methods offer complementary strategies which may lead to different insights, in this letter we highlighted several benefits of a network-specific approach. i) Ensemble methods cannot account for variability across networks within the ensemble (see fig. 2 and ref. [18]). ii) Application of any ensemble approach to a real-world network requires *a priori* assumptions about that particular network (*e.g.*, that it either lacks complexity not accounted for in the ensemble or that its effects are small). iii) Our approach naturally accounts for degree correlations of the non-Markovian type. iv) Arbitrary targeting for node and/or link removal can be easily handled with our approach. v) A network-specific analysis allows one to study deterministic attacks. Our results help explain why strategically decreasing and increasing  $\lambda$  offer fundamental strategies to respectively attack [5] and protect [7] networks.

We have provided several examples showing that our results are robust to various attack strategies, degree distributions, degree correlations, and moderate clustering. However, when strong community structure is added we expect our analysis to break down as occurs with the ensemble approaches [27]. Finding a network-specific theory accounting for communities remains open for future research. Another direction of future work includes applying our techniques to more complicated percolation problems such as  $k$ -core percolation and Achlioptas processes. However, it is likely that efforts following the techniques provided here will also be restricted to the subcritical regime.

\*\*\*

We thank the NSF for financial support through Grant No. DMS-0908221.

## REFERENCES

- [1] ALLARD A., NOËL P.-A. and DUBÉ L.J., *Phys. Rev. E*, **79** (2009) 036113.
- [2] LARREMORE D. B., SHEW W. L. and RESTREPO J. G., *Phys. Rev. Lett.*, **106** (2011) 058101; LARREMORE D. B. *et al.*, *Chaos*, **21** (2011) 025117.
- [3] MOTTER A. and LAI Y.-C., *Phys. Rev. E*, **66** (2002) 065102; MORENO Y. *et al.*, *Europhys. Lett.*, **62** (2003) 292; MOTTER A., *Phys. Rev. Lett.*, **93** (2004) 098701.
- [4] GALLOS L. K. *et al.*, *Phys. Rev. Lett.*, **94** (2005) 188701.
- [5] RESTREPO J. G., OTT E. and HUNT B., *Phys. Rev. Lett.*, **97** (2006) 094102.
- [6] REUVEN R., HAVLIN S. and BEN-AVRAHAM D., *Phys. Rev. Lett.*, **91** (2003) 247901.
- [7] MILANESE A., SUN J. and NISHIKAWA T., *Phys. Rev. E*, **81** (2010) 046112; TAYLOR D. and RESTREPO J. G., *Phys. Rev. E*, **83** (2011) 066112.
- [8] SCHNEIDER C. M. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **108** (2011) 3838.
- [9] COHEN R. *et al.*, *Phys. Rev. Lett.*, **85** (2000) 4626; CALLAWAY D. S. *et al.*, *Phys. Rev. Lett.*, **85** (2000) 5468.
- [10] NEWMAN M. E. J., STROGATZ S. H. and WATTS D. J., *Phys. Rev. E*, **64** (2001) 026118.
- [11] BOGUÑA M. and SERRANO M. Á., *Phys. Rev. E*, **72** (2005) 016106.
- [12] NEWMAN M. E. J., *Phys. Rev. Lett.*, **89** (2002) 208701.
- [13] RESTREPO J. G., OTT E. and HUNT B. R., *Phys. Rev. Lett.*, **100** (2008) 058701.
- [14] BEKESSY A., BEKESSY P. and KOMLOS J., *Stud. Sci. Math. Hung.*, **7** (1972) 343; BENDER E. A. and CANELD E. R., *J. Comb. Theory, Ser. A*, **24** (1978) 296.
- [15] CHUNG F. and LU L., *Ann. Comb.*, **6** (2002) 125.
- [16] HOLLAND P. W. and LEINHARDT S., *Comp. Group Stud.*, **2** (1971) 107; WATTS D. J. and STROGATZ S. H., *Nature*, **393** (1998) 440.
- [17] GIRVAN M. and NEWMAN M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 7821; CLAUSET A., NEWMAN M. E. J. and MOORE C., *Phys. Rev. E*, **70** (2004) 066111.
- [18] CARLSON N., KIM D.-H. and MOTTER A. E., *Chaos*, **21** (2011) 025105.
- [19] TSAPARAS P., available at: [www.cs.helsinki.fi/u/tsaparas/MACN2006/data-code.html](http://www.cs.helsinki.fi/u/tsaparas/MACN2006/data-code.html) (2006).
- [20] BATAGELJ V. and MRVAR A., available at: <http://vlado.fmf.uni-lj.si/pub/networks/data/>.
- [21] MELNIK S. *et al.*, *Phys. Rev. E*, **83** (2011) 036112.
- [22] GOLUB G. H. and VAN LOAN C. F., *Matrix Computations*, 3rd edition (Johns Hopkins University Press) 1999.
- [23] ERDÖS P. and RÉNYI A., *Publ. Math. Hung. Acad. Sci.*, **5** (1960) 17.
- [24] NELSON D. L., MCENVOY C. L. and SCHREIBER T. A., downloaded from [www.usf.edu/FreeAssociation/](http://www.usf.edu/FreeAssociation/) (1998).
- [25] TRAUD A. L. *et al.*, arXiv:0809.0690 (2010).
- [26] RIPEANU M., FOSTER I. and IAMNITCHI A., *IEEE Internet Comput.*, **6** (2002) 50, available at: [snap.stanford.edu/data/p2p-Gnutella08.html](http://snap.stanford.edu/data/p2p-Gnutella08.html).
- [27] BAGROW J., LEHMANN S. and AHN Y.-Y., arXiv:1102.5085 (2011).