

Numerical assessment of the percolation threshold using complement networks

Giacomo Rapisardi,^{1,*} Guido Caldarelli,^{1,2} and Giulio Cimini^{1,2}

¹*IMT School for Advanced Studies, 55100 - Lucca, Italy*

²*Istituto dei Sistemi Complessi (ISC)-CNR, 00185 - Rome, Italy*

(Dated: December 5, 2018)

Models of percolation processes on networks currently assume locally tree-like structures at low densities, and are derived exactly only in the thermodynamic limit. Finite size effects and the presence of short loops in real systems however cause a deviation between the empirical percolation threshold p_c and its model-predicted value π_c . Here we show the existence of an empirical linear relation between p_c and π_c across a large number of real and model networks. Such a putatively universal relation can then be used to correct the estimated value of π_c . We further show how to obtain a more precise relation using the concept of the complement graph, by investigating on the connection between the percolation threshold of a network, p_c , and that of its complement, \bar{p}_c .

Percolation theory is a widely used concept in statistical physics [1], in particular in the field of complex networks to study critical phenomena, resilience and spreading processes [2–4]. However, percolation properties in network models (be they sparse treelike graphs [5] or clustered networks [6, 7]) are often considerably different from those of real-world networks—which feature a highly more complex topology. Recently, percolation has been reformulated as a message passing process which takes as input the detailed topology of a given network to predict percolation-related observables [8, 9], and which implies that the bond percolation threshold π_c of the network is bounded from below by the leading eigenvalue of its non-backtracking matrix [10]. This approach has been then generalized to clustered networks [11], in order to go beyond the locally treelike approximation which is not reliable for networks with high density of edges and short loops [12]. However, the method comes at a price of much higher computational complexity, and is not particularly satisfactory for bond percolation processes. Another important aspect of the message passing approach is that it describes network percolation in the thermodynamic limit, and as such cannot be precisely applied to finite graphs [13]. Indeed, the very percolation transition is ill defined in finite systems.

Numerical simulations of the percolation process obtain the value of the percolation threshold p_c using Monte Carlo techniques [14]. Given Q independent realizations of the process at fixed percolation probability p , and the relative size of the largest cluster in the network $s_q(p)$, $q = 1, \dots, Q$ in the q -th realization, the percolation strength at p is estimated as $S(p) = \frac{1}{Q} \sum_q s_q(p)$, and the susceptibility as $\chi(p) = \frac{1}{Q \cdot \bar{S}(p)} \sum_q [s_q(p)]^2 - S(p)$. The best estimate of the percolation threshold is then the value of p at which the susceptibility is maximal. As the simulated system is finite, such defined pseudo-critical threshold $p_c(N)$ decays towards the percolation threshold as $p_c(N) - p_c \sim N^{-1/\nu}$, where N is the (finite) size of the network [15].

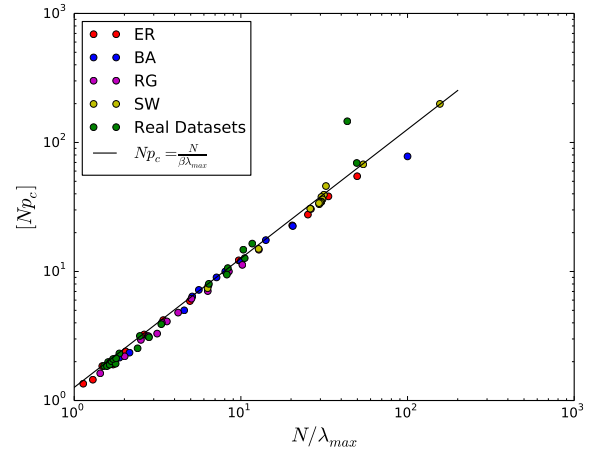


FIG. 1. Plot of Np_c versus $N\pi_c = N/\lambda_{max}$, where λ_{max} is the leading eigenvalue of the non-backtracking matrix, for several model and empirical networks. Note that accounting for the factor N allows to compare networks of different size. The black solid line is the linear fit $p_c = \pi_c/\beta$.

Figure 1 shows, for the bond percolation problem, the relation between the value p_c obtained in numerical simulations and the theoretical π_c given by the inverse of the leading eigenvalue of the non-backtracking matrix (λ_{max}). The plot is obtained by considering a total of 79 networks of different sizes N (varying approximately from 20 to 890), 23 of which are empirical while the remaining 56 are artificially generated according to four different random network models: Erdős-Rényi (ER), Regular (RG), Barabási-Albert (BA) and Watts-Strogatz (SW) [16]. Points are well fitted by a linear relation $p_c = \pi_c/\beta$ with $\chi^2/\nu = 4.34$, where however the value of $\beta = 0.791 \pm 0.019$ is different from unity: numerical and theoretical percolation thresholds do not coincide, yet their ratio appears to be constant across a variety of empirical and model networks of different size. While assessing the general validity of such an empirical evidence needs further statistical analysis, this relation can be quite valuable for correcting the theoretical value of

* giacomo.rapisardi@imtlucca.it

π_c for finite, non-treelike networks.

In this work we explore the possibility to improve such an empirical relation using the concept of *complement graph*. The complement of a graph G is the graph \bar{G} with the same vertex set, but whose edges are those which are not present in G [17, 18]. The union graph of G and \bar{G} is therefore a complete graph. Complement graphs are found since long in the mathematic literature, for instance to address the graph coloring problem [19], to develop graph compression schemes [20] and search algorithms [21], to study network synchronizability [22], to assess graph hyperbolicity [23] and domination numbers [24]. The common approach of these studies is to prove rigorous results for graphs with a small number of vertices [25–27]. Here, for the first time to our knowledge, we use complement graphs in the context of percolation on large-scale complex networks. In particular, we investigate on the existence of a complement relation for the percolation threshold p_c of a given graph G and the complement percolation threshold \bar{p}_c of \bar{G} .

Now, since the complement of a sparse network is dense, in the thermodynamic limit the percolation threshold of \bar{G} converges to the inverse of the leading eigenvalue of the adjacency matrix of \bar{G} [28]. In the simple case of ER networks, for $N \rightarrow \infty$ it is $p_c \simeq 1/\langle k \rangle = 1/[(N-1)f]$ (where f is the probability of existence of an edge), and thus the following relation should hold:

$$\frac{1}{(N-1)p_c} + \frac{1}{(N-1)\bar{p}_c} \simeq 1 \quad (1)$$

(an analogous complement relation of the two critical points also holds for regular graphs). As Figure 2 shows, eq. (1) slightly overestimates the relation between p_c and \bar{p}_c , as they do not add up to unity. In particular, the theoretical curve seems to constitute a boundary in the (p_c, \bar{p}_c) plane, and data are better fitted by a shifted linear relation

$$\frac{1}{(N-1)p_c} + \frac{1}{(N-1)\bar{p}_c} = \alpha < 1, \quad (2)$$

with $\alpha = 0.889 \pm 0.008$ and $\chi^2/\nu = 3.68$. The same behavior is observed in Figure 3 for theoretical values of the percolation threshold, obtained as the leading eigenvalue of the non-backtracking matrices.

Building on the analysis of Figure 1, we now study the relation

$$p_c + \bar{p}_c = \frac{1}{\beta'} \left(\frac{1}{\lambda_{max}} + \frac{1}{\bar{\lambda}_{max}} \right). \quad (3)$$

As shown in Figure 4, eqn.(3) fits the data quite well, and much better than the fit of Figure 1. From the fit we obtained $\beta' = 0.856 \pm 0.010$ and $\chi^2/\nu = 1.16$. This factor can therefore be used to improve the estimate of the percolation threshold on finite, non treelike networks.

To show that this is the case, in Figure (5) we compare different estimates of the numerical percolation threshold, obtained as either the leading eigenvalues of the adjacency matrix λ_{max}^A or of the non-backtracking matrix

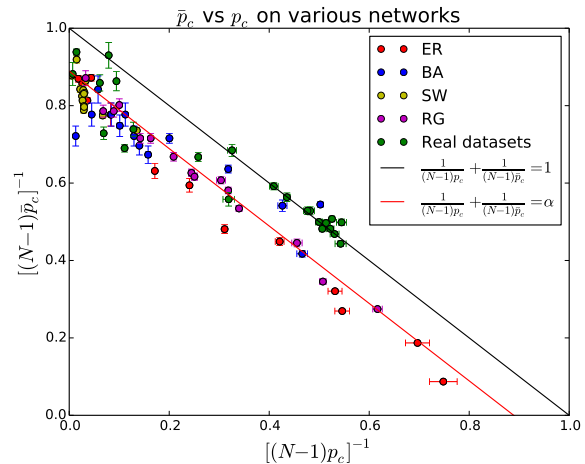


FIG. 2. Plot of p_c^{-1} versus \bar{p}_c^{-1} for various networks of different sizes. The solid black line is eq. (1), while the solid red line is the linear fit of data.

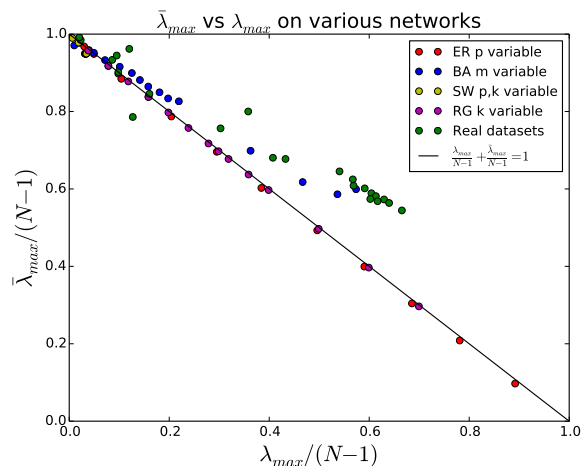


FIG. 3. Plot of λ_{max} versus $\bar{\lambda}_{max}$ for various networks of different sizes.

λ_{max}^{NB} , eventually corrected by the β' factor. We indeed see that β' can be used to improve, on average, the approximation given by theoretical models.

We believe that the corrective factor is related to finite size effects and non treelike structures, however this hypothesis needs further investigation. Overall, while our approach is just at infant stage and our findings are only preliminary, they may have important concrete applications.

Acknowledgments

This work was supported by the EU projects CoeGSS (grant n. 676547) and SoBigData (grant n. 654024).

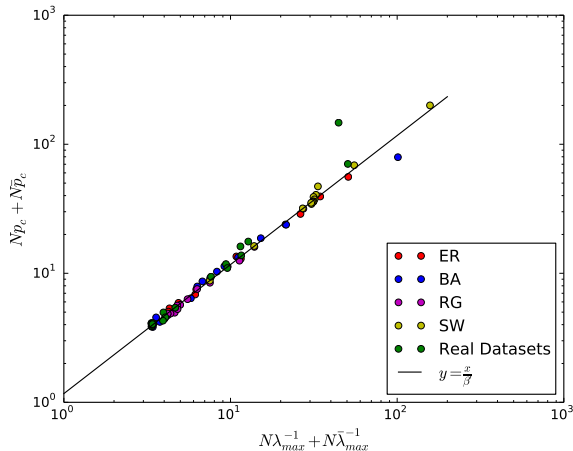


FIG. 4. Plot of $Np_c + N\bar{p}_c$ vs $\left(\frac{N}{\lambda_{max}} + \frac{N}{\lambda_{max}^*}\right)$, where λ_{max} is the leading eigenvalue of the non-backtracking matrix, for several model and empirical networks. The black solid line is the linear fit of eq. (3).

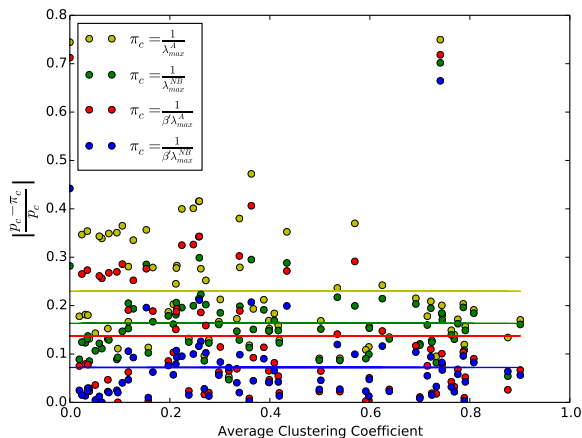


FIG. 5. Relative error associated to different estimates for the percolation threshold. Solid lines indicates average values of dots of the same color.

Appendix: definition and properties of the complement network

Formally, let a_{ij} be the generic element of the adjacency matrix A associated with a given binary undirected graph G of N vertices, such that $a_{ij} = 1$ if an edge between vertices i and j exists, and $a_{ij} = 0$ otherwise. The adjacency matrix of the complement graph \bar{G} is defined through $\bar{a}_{ij} = 1 - \delta_{ij} - a_{ij}$, where δ_{ij} is the Kronecker delta which excludes self loops from \bar{G} , and $1 - \delta_{ij}$ defines the adjacency matrix of the complete graph. It follows trivially that $M + \bar{M} = \binom{N}{2}$, $\rho + \bar{\rho} = 1$ and $k_i + \bar{k}_i = N - 1 \forall i$, where M , ρ and k_i denote the number of edges, the edge density, and the degree of (number of edges incident with) generic vertex i , respectively. Thus, given the degree distribution $P(k)$, the distribution of the complement degree is obtained as $\bar{P}(\bar{k}) = P(N - 1 - \bar{k})$, *i.e.*, as the reflection of $P(k)$ on the $\frac{N-1}{2}$ vertical axis. Notably, the degree distribution of both a regular graph and an Erdős-Rényi graph (ER) are invariant under this transformation: the complement of a regular graph is a regular graph, as the complement of an ER is an ER. In particular, the complement of an ER with connection probability f is an ER with connection probability $1 - f$.

Moving to higher-order properties, the number of triangles (closed loop of length 3) of a graph and of its complement is

$$\Sigma_{\Delta} = \frac{\text{Tr}A^3 + \text{Tr}\bar{A}^3}{6} = \binom{N}{3} - \frac{1}{2} \sum_i k_i \bar{k}_i. \quad (4)$$

As such, both cases $k_i = 0$ and $\bar{k}_i = 0 \forall i$ (empty and complete graph) lead to $\Sigma_{\Delta} = \binom{N}{3}$ as expected. As for transitivity, a complementarity relation can be written also for the local clustering coefficient $c_i = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ij} a_{jk} a_{ik}}{k_i(k_i - 1)}$:

$$c_i k_i (k_i - 1) + \bar{c}_i \bar{k}_i (\bar{k}_i - 1) = k_i^{nn} k_i + \bar{k}_i^{nn} \bar{k}_i - k_i - \bar{k}_i - k_i \bar{k}_i, \quad (5)$$

where $k_i^{nn} = \frac{\sum_{j \neq i} a_{ij} k_j}{k_i}$ is the average nearest-neighbors degree.

[1] D. Stauffer and A. Aharony, *Introduction To Percolation Theory* (Taylor & Francis, 1994).
 [2] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Physical Review Letters* **85**, 5468 (2000).
 [3] M. E. J. Newman, *Physical Review E* **66**, 016128 (2002).
 [4] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Review of Modern Physics* **80**, 1275 (2008).
 [5] R. Cohen, D. ben Avraham, and S. Havlin, *Physical Review E* **66**, 036113 (2002).
 [6] M. A. Serrano and M. Boguñá, *Physical Review Letters* **97**, 088701 (2006).
 [7] M. E. J. Newman, *Physical Review Letters* **103**, 058701 (2009).
 [8] B. Karrer, M. E. J. Newman, and L. Zdeborová, *Physical*

Review Letters **113**, 208702 (2014).
 [9] K. E. Hamilton and L. P. Pryadko, *Physical Review Letters* **113**, 208701 (2014).
 [10] K. Hashimoto and Y. Namikawa, *Automorphic Forms and Geometry of Arithmetic Varieties*, Advanced Studies in Pure Mathematics (Elsevier Science, 2014).
 [11] F. Radicchi and C. Castellano, *Physical Review E* **93**, 030302 (2016).
 [12] F. Radicchi, *Physical Review E* **91**, 010801 (2015).
 [13] G. Timár, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, *Physical Review E* **95**, 042322 (2017).
 [14] M. E. J. Newman and R. M. Ziff, *Physical Review Letters* **85**, 4104 (2000).
 [15] F. Radicchi and C. Castellano, *Nature Communications*

- 6**, 10196 (2015).
- [16] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
- [17] L. Clark and R. Entringer, *Periodica Mathematica Hungarica* **14**, 57 (1983).
- [18] J. Gross and J. Yellen, *Graph Theory and Its Applications*, *Discrete Mathematics and Its Applications* (Taylor & Francis, 1998).
- [19] E. A. Nordhaus and J. W. Gaddum, *The American Mathematical Monthly* **63**, 175 (1956).
- [20] M.-Y. Kao, N. Occhiogrosso, and S.-H. Teng, *Journal of Combinatorial Optimization* **2**, 351 (1998).
- [21] H. Ito and M. Yokoyama, *Information Processing Letters* **66**, 209 (1998).
- [22] Z. Duan, C. Liu, and G. Chen, *Physica D: Nonlinear Phenomena* **237**, 1006 (2008).
- [23] S. Bermudo, J. M. Rodriguez, J. M. Sigarreta, and E. Tours, *Applied Mathematics Letters* **24**, 1882 (2011).
- [24] R. Haas and T. B. Wexler, *Discrete Mathematics* **283**, 87 (2004).
- [25] J. Akiyama and F. Harary, *International Journal of Mathematics and Mathematical Sciences* **2**, 223 (1979).
- [26] S. Xu, *Discrete Mathematics* **65**, 197 (1987).
- [27] M. Petrović, Z. Radosavljević, and S. Simić, *Linear and Multilinear Algebra* **51**, 405 (2003).
- [28] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan, *The Annals of Probability* **38**, 150 (2010).