Effect of degree correlations above the first shell on the percolation transition

Effect of degree correlations above the first shell on the percolation transition

L. D. Valdez[1(a)], C. Buono[1], L. A. Braunstein[1,2] and P. A. Macri[1]

[1] *Instituto de Investigaciones Físicas de Mar del Plata (IFIMAR)-Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata-CONICET - Funes 3350, (7600) Mar del Plata, Argentina*
[2] *Center for Polymer Studies, Boston University - Boston, MA 02215, USA*

Abstract – The use of degree-degree correlations to model realistic networks which are characterized by their Pearson's coefficient, has become widespread. However the effect on how different correlation algorithms produce different results on processes on top of them, has not yet been discussed. In this letter, using different correlation algorithms to generate assortative networks, we show that for very assortative networks the behavior of the main observables in percolation processes depends on the algorithm used to build the network. The different alghoritms used here introduce different inner structures that are missed in Pearson's coefficient. We explain the different behaviors through a generalization of Pearson's coefficient that allows to study the correlations at chemical distances ℓ from a root node. We apply our findings to real networks.

Introduction. – In the last two decades, the use of complex networks in the study of many processes, such as the spread of diseases, random or intentional attacks, synchronization, etc. [1–3], has led to a deeper understanding of these processes. For example, the use of network-based models in epidemiology has demonstrated how the topology affects the total fraction of the infected population and how its knowledge can be used to develop efficient immunization strategies [4]. Each of the aspects of the network topology has a substantial effect on the processes taking place on the underlying network.

A first feature of the network topology is the degree distribution $P_d(k)$, *i.e.*, the fraction of nodes with degree k. One of the most used networks is the Erdös-Rényi (ER) [5] with $P_d(k) = e^{-\langle k \rangle} \langle k \rangle^k / k!$, where $\langle k \rangle$ is the average degree. However, in real networks the degree distribution fits better to a scale-free (SF) $P_d(k) \sim k^{-\lambda}$, where λ is the broadness of the distribution. Many of these networks also have an exponential cutoff κ [6], and the degree distribution can be represented by $P_d(k) \sim k^{-\lambda} e^{-k/\kappa}$.

Several analytical and numerical models on degree-degree uncorrelated networks were developed with the

above distributions [7,8]. However, in the last years it has been observed that in real networks, the degrees of the linked nodes are correlated. For example, it is known that in social networks, nodes tend to be linked to others with similar degree, while in technological networks, large degree nodes tend to be connected with low-degree ones. Networks with the first pattern of connections are called assortative, while those with the second pattern are disassortative. Recent researches suggest that networks adopt a correlated structure in order to optimize some processes that are developed on top of them, such as synchronization [9], transport [10], traffic dynamic [11], congestion [12] and growth [13]. As a consequence, it is necessary to consider models of networks with degree-degree correlations in order to study how the processes are affected by them. There are different measures to quantify the correlations between nearest neighbors [1,2]. The most detailed measure is the joint degree distribution $P_e(k, k')$, which indicates the probability that a node of degree k is linked with another of degree k'. However, due to the large amount of information contained in this distribution, it is difficult to interpret the results based on $P_e(k, k')$. Hence more global measures are needed to simplify the interpretation of the results. For example,

_____
(a)E-mail: ldvaldes@mdp.edu.ar

a measure which follows from the above and is more simplified is the average nearest-neighbor degree of a node with connectivity k, $k_{nn}(k) = \sum_{k'} k' P_e(k|k')$. The sign of the slope of $k_{nn}(k)$ is positive (negative) for assortative (disassortative) networks. However, the most used measurement is Pearson's coefficient, which is less detailed than the previous ones and gives a general overview of the correlation to nearest neighbors on the network. Pearson's coefficient is given by

$$r = \frac{\langle kk' \rangle_e - \langle (k+k')/2 \rangle_e^2}{\langle (k^2+k'^2)/2 \rangle_e - \langle (k+k')/2 \rangle_e^2}, \qquad (1)$$

where e is the average over $P_e(k,k')$. Assortative networks have $r > 0$, while disassortative networks have $r < 0$. Finally, networks with $r = 0$ are called uncorrelated.

Many studies that concentrate on how the degree-degree correlation affects the processes on top of networks, use Pearson's coefficient to quantify the results. For example in ref. [9] it has been observed that synchronization is enhanced for $r \neq 0$. Also, in ref. [12], under a gradient network formalism, the authors explain why in disassortative networks the congestion is lower than in uncorrelated and assortative ones. On the other hand, some studies on the effect of degree-degree correlations were focused on the link percolation process [14–17]. The investigations on this process are very useful since they provide information about how resistant a network is against random failures and also due to the relation between percolation with some aspects of the spread of diseases [18–22]. For uncorrelated networks, it is well known that in percolation processes there is a critical fraction $p = p_c$ of nodes/links, above which the network undergoes a second-order phase transition [7,19]. The order parameter of the transition is $P_\infty(p)$, which is the fraction of nodes that belong to the largest component for a given value of p. For infinite systems ($N \rightarrow \infty$) and close to the threshold (p_c), the order parameter behaves as [23]

$$P_\infty(p) \sim (p-p_c)^\beta, \quad p \gtrsim p_c + \delta p, \quad \delta p > 0, \qquad (2)$$

and the average value of the size of the finite clusters $\langle s(p) \rangle$ goes as

$$\langle s(p) \rangle \sim (p-p_c)^{-\gamma}, \quad p \lesssim p_c - \delta p, \quad \delta p > 0, \qquad (3)$$

where $\beta = 1$ and $\gamma = 1$ in the mean-field approach [23]. In the case of finite networks with N nodes, eq. (3) obeys the general scaling law

$$\langle s \rangle \equiv \langle s(p,N) \rangle = N^{\gamma \Theta} f((p-p_c(N))N^\Theta), \qquad (4)$$

where f is a scaling function that behaves as a constant at the threshold, with $\langle s \rangle \sim N^{1/3}$ for uncorrelated ER and SF networks with $\lambda \geqslant 4$ [24,25].

Newman [26] and Miller [27] showed that networks with $r > 0$ are more resilient to damage than those with $r \leqslant 0$. However, there has been little discussion in the literature about whether the percolation phase transition remains of

second order depending on the value of r. This issue was only addressed in ref. [28] where they found that in correlated networks constructed using a Metropolis algorithm and characterized by r, for the disassortative case percolation belongs to the same universality class as uncorrelated networks but highly assortative networks have no second-order phase transition. A recent investigation by Dorogovtsev *et al.* [29] pointed out that recursive trees can be constructed with a highly correlated structure but with zero Pearson's coefficient. On the other hand, in ref. [30] the authors show that positive correlations affects the loop statistics on networks. Thus, Pearson's coefficient could hide a more complex pattern of connections between nodes that also affect the processes that spread on networks. This raises the question as to whether or not Pearson's coefficient is a reliable tool for characterizing correlated networks and, in particular, the percolation transition on assortative networks.

In this letter we use different correlation algorithms to determine the effect of very assortative networks on the link percolation transition. We find that the results depend strongly on the algorithm used to correlate the networks and not on the value of Pearson's coefficient. Moreover, we find that strong correlations above the first shell on assortative networks affect the percolation transition.

Correlation algorithms. – We will focus only on assortative networks because we want to understand the effect of different algorithms in the percolation transition.

In this work we use rewiring algorithms to correlate the networks, since they preserve the degree distribution $P_d(k)$. In particular we focus on two correlation algorithms: the exponential random graph (ERG) [28] and the local optimal algorithm (LOA) introduced by us.

For the ERG algorithm, the process to correlate a network or graph G [28], is based on a Metropolis dynamic, which uses a "Hamiltonian" function given by $H(G) = -J \sum_{i,j>i} A_{ij} k_i k_j$, where A_{ij} is the component of the adjacency matrix, k_i (k_j) is the connectivity of node i (j) and J is a control parameter for the correlation. For $J > 0$ ($J < 0$) an assortative (disassortative) network is built. This algorithm generates a Gibbsian ensemble network and is ergodic. The algorithm correlates the network, successively applying the following steps:

1) in a network G, two links are chosen randomly to be correlated;

2) rewiring is allowed (from configuration G to G') with probability $w = \min\{1, \exp[-(H(G') - H(G))]\}$, disallowing autoloops and multiple connections.

The rewiring process continues until the steady state, where for large enough system sizes, r depends only on J (not shown here).

In the LOA algorithm, a network is correlated as follows:

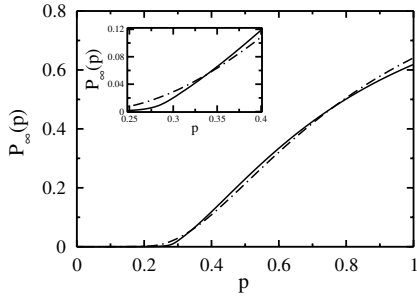1) two links are randomly chosen from the network to be correlated;

Fig. 1: Plot of $P_\infty(p)$ for ER networks with $\langle k \rangle = 2$, $N = 8 \times 10^4$ and $r = 0.78$, correlated with LOA (solid line) and ERG (dashed line, $J = 1$). In the inset we show an enlargement of the main plot close to the threshold. The simulations were done over 10^3 realizations.

2) for assortative (disassortative) networks, we choose the configuration that increases (decreases) most Pearson's coefficient r without generating autoloops or multiple connections.

This algorithm is not ergodic but achieves the desired correlation faster, since in contrast to the previous one, we do not have to wait for the equilibrium value in order to reach the desired value of r.

Results and discussion. –

Link percolation on assortative networks. In this section we study the effects of the correlation algorithms on the percolation transition in assortative networks.

In fig. 1, we show $P_\infty(p)$ for an ER network with fixed N and r for the ERG and LOA models. Even though the curves do not differ significantly between them, in the inset we can see that $P_\infty(p)$, close to the threshold computed with different algorithms converge to zero with different slopes, suggesting that the transitions are different even if the same value of r is used in both algorithms. The behavior of $P_\infty(p)$ for networks generated with the ERG algorithm seems to be nonsingular.

It is well known that in the thermodynamic limit a second-order phase transition, has a singularity in the derivative of the order parameter at the threshold p_c. In fig. 2 we plot the derivative of $P_\infty(p)$ with respect to p for different networks sizes. From the plots, we can see that in strong positive correlated networks with the LOA algorithm, the derivative of $P_\infty(p)$ exhibits a singularity as we increase the system size, characteristic of a second-order phase transition. However, for the ERG model, the slope is smooth and does not depend on N. Therefore, we confirm that the transitions for the ERG and the LOA are not the same [28].

In order to verify the existence of a second-order phase transition, we compute $\langle s \rangle$ for ER and different network sizes as shown in fig. 3. We can see that with the LOA correlation model as N increases, the peak of $\langle s \rangle$ increases around the critical threshold $p_c(N)$ as in a
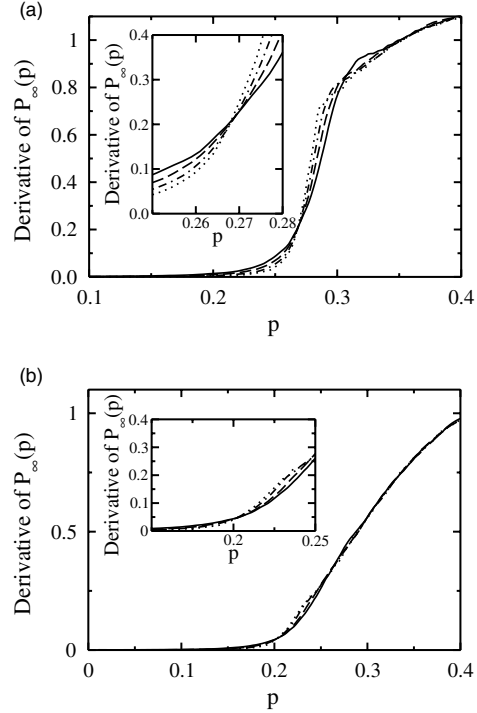

Fig. 2: Plot of the derivative of P_∞ as a function of p for ER with $\langle k \rangle = 2$, $r = 0.78$, correlated with LOA (a) and ERG (b) for $N = 8 \times 10^4$ (solid line), $N = 16 \times 10^4$ (dashed line), $N = 32 \times 10^4$ (dash-dotted line) and $N = 64 \times 10^4$ (dotted line). In the insets we show an enlargement of the main plots close to the threshold. The simulations were done over 10^3 realizations.

second-order phase transition [23]. At $p = p_c$ (see eq. (4)) $\langle s \rangle \sim N^{0.16}$, suggesting that percolation in assortative networks generated by the LOA belongs to a different universality class from that in uncorrelated networks. As noted by Noh using the ERG model [28], $\langle s \rangle$ has no peak independent of the value of N, which still indicates that the transition is not of second order.

We run all our simulations for SF networks with exponential cutoff and find similar behavior that for ER networks. In fig. 4 we show that the derivative of $P_\infty(p)$ and $\langle s \rangle$ do not exhibit any singularity as we increase the network size for the ERG model, in contrast to the LOA model.

Despite the discrepancies between the assortative networks, we find that disassortative networks generated by both algorithms undergo a second-order phase transition with $\langle s \rangle \sim N^{1/3}$, which not only confirms that disassortative networks belong to the same universality class as uncorrelated ones, but also that the critical exponents are independent of the algorithm used to generate negative correlations. Consequently, we conclude that while Pearson's coefficient is a good measure to uniquely characterize the percolation transition on disassortatives
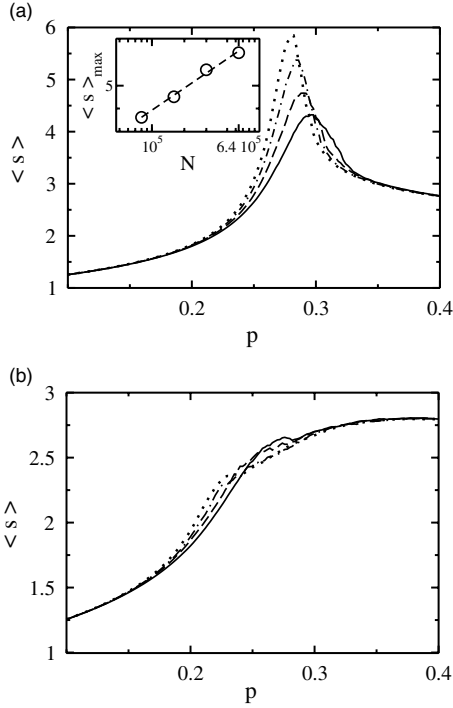
Fig. 3: Plot of $\langle s \rangle$ as a function of p for the same networks as in fig. 2: LOA (a) and ERG (b). In the inset of (a) we plot the height of the peak of $\langle s \rangle$ as a function of N on a log-log scale. The dashed line represents a power law fitting from which we obtain $\gamma\Theta = 0.16$ (see eq. (4)).
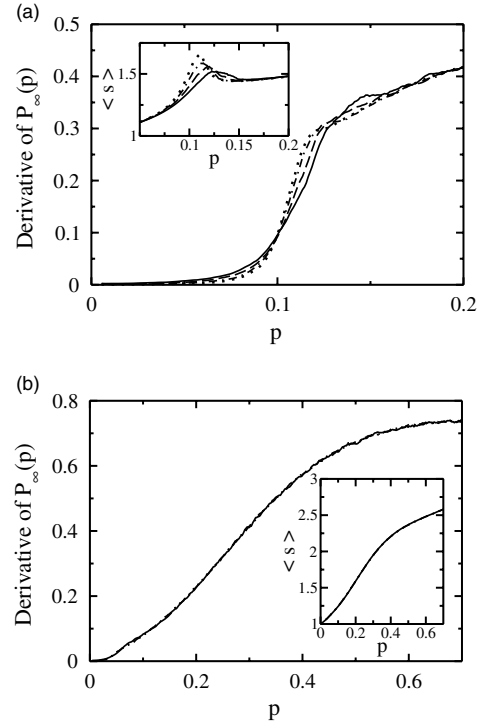

Fig. 4: Derivative of $P_\infty(p)$ for SF networks with $\lambda = 2$, $\kappa = 10$ and $r = 0.55$ correlated with LOA (a) and ERG (b) for $N = 8 \times 10^4$ (solid line), $N = 16 \times 10^4$ (dashed line), $N = 32 \times 10^4$ (dash-dotted line) and $N = 64 \times 10^4$ (dotted line). In the inset we plot $\langle s \rangle$ as a function of p for LOA (a) and ERG (b).
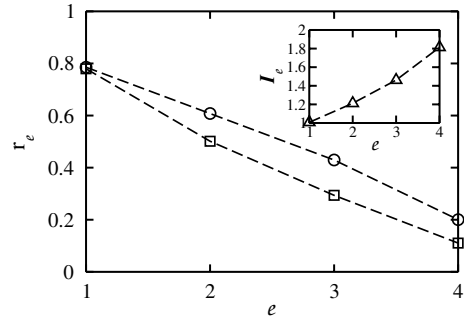
and uncorrelated networks, it fails to characterize the transition on very assortative networks. The discrepancies between the assortative networks generated with different algorithms must underlie on structural differences introduced by the algorithms, which cannot be explained only by Pearson's coefficient. Hence, we will show below the different effects that both algorithms introduce on the network topology.

Correlations above the first shell. In order to explain the differences between assortative networks with the same value of r but correlated with different algorithms, we propose a new measure called generalized Pearson's coefficient r_ℓ, which measures the correlation between the connectivity of a root node and the neighbors at a chemical distance ℓ, *i.e.* which are located in the ℓ-th shell from the root. In this way, we denote the usual Pearson's coefficient r as r_1, Pearson's coefficient to second neighbors as r_2 and so forth. We define,

$$r_\ell = \frac{\langle kk' \rangle_\ell - \langle (k+k')/2 \rangle_\ell^2}{\langle (k^2 + k'^2)/2 \rangle_\ell - \langle (k+k')/2 \rangle_\ell^2}, \tag{5}$$

where $\langle \ldots \rangle_\ell$ is the average over $P_\ell(k, k'_\ell)$, where $P_\ell(k, k'_\ell)$ is the probability that a node of degree k is connected to a node of degree k' at the ℓ-th shell.

Fig. 5: Plot of r_ℓ as a function of the shell number ℓ for LOA ($\square$) and ERG ($\bigcirc$) in ER networks with $\langle k \rangle = 2$, $r = r_1 = 0.78$ and $N = 10^4$. The dashed lines are used only as a guide to the eye. In the inset we show I_ℓ as a function of the shell number ℓ. As seen in the plot $I_2 = 1.21$, $I_3 = 1.46$, $I_4 = 1.81$. Then the ERG model generates stronger long-range correlations than the LOA model. The simulations were done over 10^4 realizations.

In fig. 5 we show for ER networks r_ℓ as a function of ℓ in assortative networks[1], for $r_\ell > 0$. We can see that r_ℓ is a decreasing function of ℓ. Imposing high assortativity in the network, correlations above the first shell from a root node

---

[1]The understanding of the disassortative behavior of r_ℓ goes beyond the scope of this work.
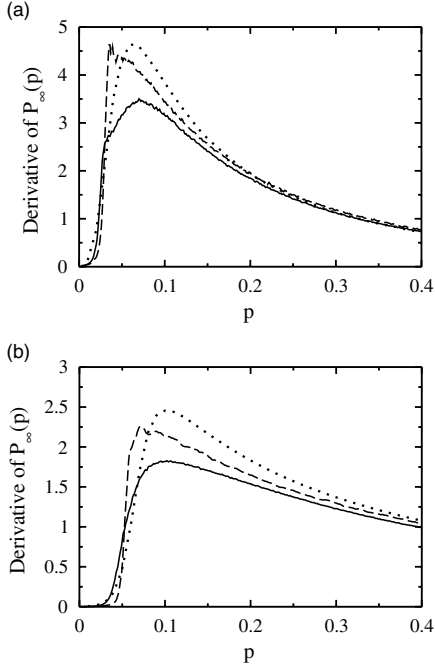
(a)

(b)

Fig. 6: Plot of the derivative of P_∞ as a function of p close to the threshold for (a) the cond-mat coauthorship network with $N = 40421$ and $r = 0.18$ and (b) the mathematics coauthorship network with $N = 391529$ and $r = 0.12$. The solid lines correspond to percolation on the original data, and the dashed and dotted lines correspond to LOA and ERG networks, respectively, with the same degree distribution as the real data. We can see that all the curves exhibit a pronounced steep growth at the threshold characteristic of a second-order phase transition.

are generated. Thus, correlation will build groups of nodes connected with similar degrees at shells close to the root, forming groups of nodes of low degrees and others with high degree. The high-degree groups will form a strong core very resilient to random failures. The stronger the correlations are above the first shell, the more resilient groups are, smoothing out the transition.

In the inset of fig. 5 we plot the factor I_ℓ defined as the ratio between r_ℓ for ERG and r_ℓ for LOA networks. Then $I_\ell > 1$ $(I_\ell < 1)$ indicates that ERG networks have higher (lower) degree correlation at a distance ℓ than LOA networks. It is easy to see that for shells above $\ell = 1$, I_ℓ increases with ℓ reaching a value of $I_4 = 1.8$ for the fourth shell[2] for $N = 10^4$. Thus, ERG networks generate stronger correlation above shell $\ell = 1$ than LOA networks supporting our picture that high assortative ERG networks posses a stronger structure of groups, with nodes of similar connectivities, than in the LOA model. This introduces two aspects of the network that contributes to erase the

second-order phase transition in the ERG model. First, ERG networks have more homogeneous groups of nodes with low connectivity that are easier to fragment, generating smaller clusters than in LOA assortative networks. On the other hand, in ERG networks there are also more homogeneous groups of highly connected nodes that remain in the largest component through the percolation process[3], preventing the network from fragmenting into large finite clusters, and smoothing out the curve of $P_\infty(p)$ near the threshold (see fig. 2). Consequently, these two aspects induce a greater presence of small clusters, leading to a non-diverging $\langle s \rangle$. To determine if this behavior is representative of real networks, we measure r_ℓ on two different positive correlated SF real graphs, the condensed matter (cond-mat, $\lambda \approx 1.60$ and $\kappa \approx 40$) [33] and mathematics ($\lambda \approx 1.72$ and $\kappa \approx 35$) [34] coauthorship networks. In both cases, we obtain that positive correlations are non-negative only until the second shell with $r_1 = r = 0.18$, $r_2 = 0.01$ for the cond-mat collaboration network and $r_1 = r = 0.12$, $r_2 = 0.03$ for the mathematics collaboration network. Consequently, real networks posses mainly correlation to first neighbors (r), in contrast to the stronger correlation structure above $\ell = 1$ found in the theoretical networks, particularly for ERG, and, therefore, the real graphs analyzed here have a less defined structure of groups of nodes connected with similar degrees, suggesting that there are constraints which prevent networks from evolving towards extreme correlations. This can be seen, for example, on both coauthorship networks used here. In general, coauthorship networks are made of research groups composed by senior researchers with high connectivity and young researchers generally with low connectivity. On the one hand, senior researchers from different research groups have some connections between themselves in order to increase the resources and the importance of their groups. Moreover, young researchers from the same group collaborate between themselves, because they generally work on similar projects. This explains the assortativity of the networks. On the other hand, in order to share their knowledges, senior researchers are highly connected to the young ones of their group, decreasing the assortativity. Thus, the network evolves with an assortativity correlation, but with a not too high Pearson's coefficient in order to improve the functionality of the full coauthorship network. This is why real assortative collaboration networks are not extreme assortativity correlated. Moreover, since this positive correlation decreases with ℓ as we can see in both real networks used here and in the Router graph studied in ref. [35], the correlation above the first shell is almost zero due to the small Pearson's coefficient, which leads to a critical behavior of the percolation process on these networks. In fig. 6 we plot for the real networks, the derivative of $P_\infty(p)$ for the original data and after applying the LOA and the ERG models for networks

_____

[2]For a SF network with $\lambda = 2$, $\kappa = 10$, $r = 0.55$ and $N = 10^4$, we obtain that $I_2 = 2.11$, however positive correlations only reach to the second shell.

[3]The same picture can be seen in onion networks [31,32]. However onion and assortativity are distinct properties as pointed out in [31].

with the same degree distribution. As we can see, there is a steep growth of the derivative near criticality, which is equally sharp in the LOA and ERG networks, supporting our argument that in real networks the second-order phase transition exists. In turn, we note that ERG and LOA networks with the same degree distribution and Pearson's coefficient as real networks, have also a second-order phase percolation transition. Our results suggest that Pearson's coefficient is a good indicator of the percolation transition order when the network is not strongly correlated above the first shell, but for high positive degree-degree correlations, different algorithms may generate different behaviors in percolation and, therefore, Pearson's coefficient is no longer useful to indicate the behavior of the transition. As a consequence, it becomes crucial to use measures that take into account the inner structure of the network, such as r_ℓ, that, as we show here, is a better indicator of whether a network will undergo or not a critical phase percolation transition.

In summary, we find that the second-order percolation transition on theoretical networks depends on the inner structure imposed by the algorithms used to build them, and not only on Pearson's coefficient. This means that Pearson's coefficient hides a long-range correlation that could change dramatically the behavior on top of them. We propose a new magnitude r_ℓ, which allows to explain the discrepancies between percolation transition on assortative networks generated by different correlation algorithms. For the real networks analyzed in this letter, we find that they posses mainly first-neighbors correlation, and, thus, the percolation transition is of second order.

$* * *$

REFERENCES

[1] Boccaletti S., Latora V., Moreno Y., Chavez M. and Hwang D., *Phys. Rep.*, **424** (2006) 175.
[2] Dorogovtsev S. N. and Mendes J. F. F., *Evolution of Networks* (Oxford University Press) 2003.
[3] Dorogovtsev S. N., Goltsev A. V. and Mendes J. F. F., *Rev. Mod. Phys.*, **80** (2008) 1275.
[4] Cohen R., Havlin S. and ben Avraham D., *Phys. Rev. Lett.*, **91** (2003) 247901.
[5] Erdös P. and Rényi A., *Publ. Math.*, **6** (1959) 290.
[6] Amaral L. A. N., Scala A., Barthélémy M. and Stanley H. E., *Proc. Natl. Acad. Sci. U.S.A.*, **97** (2000) 11149.
[7] Newman M. E. J., Strogatz S. H. and Watts D. J., *Phys. Rev. E*, **64** (2001) 026118.
[8] Miller J. C., *Phys. Rev. E*, **76** (2007) 010101.
[9] Sorrentino F., di Bernardo M., Cuéllar G. H. and Boccaletti S., *Physica D*, **224** (2006) 123.
[10] Xue Y.-H., Wang J., Li L., He D. and Hu B., *Phys. Rev. E*, **81** (2010) 037101.
[11] Sun J.-T., Wang S.-J., Huang Z.-G. and Wang Y.-H., *Phys. A*, **388** (2009) 3244.
[12] Pastore y Piontti A. L., Braunstein L. A. and Macri P. A., *Phys. Lett. A*, **374** (2010) 4658.
[13] Shao J., Buldyrev S. V., Braunstein L. A., Havlin S. and Stanley H. E., *Phys. Rev. E*, **80** (2009) 036105.
[14] Vázquez A. and Moreno Y., *Phys. Rev. E*, **67** (2003) 015101.
[15] Xulvi-Brunet R. and Sokolov I. M., *Phys. Rev. E*, **70** (2004) 066102.
[16] Moreira A. A., Andrade J. S., Herrmann H. J. and Indekeu J. O., *Phys. Rev. Lett.*, **102** (2009) 018701.
[17] Hooyberghs H., Van Schaeybroeck B., Moreira A. A., Andrade J. S., Herrmann H. J. and Indekeu J. O., *Phys. Rev. E*, **81** (2010) 011102.
[18] Callaway D. S., Newman M. E. J., Strogatz S. H. and Watts D. J., *Phys. Rev. Lett.*, **85** (2000) 5468.
[19] Cohen R., Erez K., ben Avraham D. and Havlin S., *Phys. Rev. Lett.*, **85** (2000) 4626.
[20] Cohen R., ben Avraham D. and Havlin S., *Phys. Rev. E*, **66** (2002) 036113.
[21] Newman M. E. J., *Phys. Rev. E*, **66** (2002) 016128.
[22] Grassberger P., *Math. Biosci.*, **63** (1983) 157.
[23] Stauffer D. and Aharony A., *Introduction to Percolation Theory* (Taylor & Francis) 1985.
[24] Bornholdt S. and Schuster H. (Editors), *Handbook of Graphs and Networks* (Wiley-VCH, Berlin) 2003.
[25] Wu Z., Lagorio C., Braunstein L. A., Cohen R., Havlin S. and Stanley H. E., *Phys. Rev. E*, **75** (2007) 066110.
[26] Newman M. E. J., *Phys. Rev. E*, **67** (2003) 026126.
[27] Miller J. C., *Phys. Rev. E*, **80** (2009) 020901.
[28] Noh J., *Phys. Rev. E*, **76** (2007) 02611.
[29] Dorogovtsev S. N., Ferreira A. L., Goltsev A. V. and Mendes J. F. F., *Phys. Rev. E*, **81** (2010) 031135.
[30] Noh J., *Eur. Phys. J. B*, **66** (2008) 251.
[31] Schneider C. M., Moreira A. A., Andrade J. S. jr., Havlin S. and Herrmann H. J., *Proc. Natl. Acad. Sci. U.S.A.*, **108** (2011) 3838.
[32] Wu Z.-X. and Holme P., *Phys. Rev. E*, **84** (2011) 026106.
[33] Newman M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **98** (2001) 404.
[34] Palla G., Farkas I. J., Pollner P., Derényi I. and Vicsek T., *New J. Phys.*, **10** (2008) 123026.
[35] Echenique P., Gómez-Gardeñes J., Moreno Y. and Vázquez A., *Phys. Rev. E*, **71** (2005) 035102.