Multiple percolation transitions in three interconnected networks

Author: Miquel Gubau Besalú
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Marián Boguñá Espinal

Abstract: The way in which complex networks behave has brought out very interesting properties. With its study, simple models have been developed to isolate these new phenomena, opening up, in turn, many possibilities. The following work aims to extend the double percolation found in the special coupling of two networks. The first step is to investigate the case of three interconnected networks. The structure chosen to be studied is the Erdös-Rényi. Afterwards, to simulate the bond percolation, I have applied the Newman-Ziff algorithm. We observed that multiple percolation transitions can occur, finding the emergence of a new discontinuous phase transition to consider.

I. INTRODUCTION

The behaviour of networks has been a largely investigated topic due to its wide applicability. For this reason, a way to deal with them needs to be established. There are some basic and interesting measurement parameters to mention. We have the network's number of nodes N and edges M. Besides, an important feature is centrality, quantifying the importance of the vertices and links. Within this classification, the degree k of a node is the number of attached edges it has. Therefore, to describe the whole system it is useful to know the network's average degree $\langle k\rangle = 2M/N$, where the links are considered double because every bond is connecting two vertices.

One of the most studied processes, thanks to its simplicity, is percolation [1, 2]. It consists of eliminating edges to see how the properties of a network change. This procedure is called bond percolation. On the other hand, the removal of the network's vertices, thus deleting the links attached to them, is known as site percolation. An important part of the process is the way in which the elimination is made, being the simplest the random one. Percolation is known to explain and be used to study many real systems like the Internet and social or epidemiological spread networks, to mention some.

Focusing on bond percolation, from a generated network where all the edges have been erased, it is useful to simulate how they are set up again. While this is done, the number of edges being occupied m characterises the system's evolution. Dividing it by the original number of links M, the bond occupation probability $p = m/M$ is obtained. This probability is the control parameter of the simulation. The remarkable point of the process is the appearance of a continuous phase transition. We say that a network percolates when the largest cluster of interconnected nodes G, also known as the largest component, becomes extensive. This means that with the system growth the largest component will grow too. Meanwhile, if the network is not percolated the largest cluster is constant even though the system grows. At this transition point, the percolation threshold p_c, it is said that a giant component is formed, or dissolved if we are deleting links.

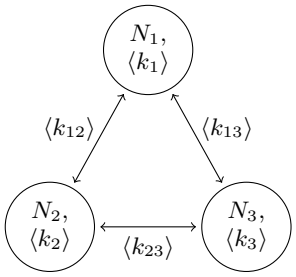

FIG. 1. General structure and parameters necessary to define the coupling of three networks.

As can be seen from the reported results in [3], the heterogeneous division of a network in two regions, a core and a periphery, leads to a double percolation phase transition. This is caused by the breaking of the same symmetry, in opposition to the modern theory of continuous phase transitions. First is observed the percolation of the core, the most connected region, followed by the whole network, including the periphery. This is because the links between regions show a sublinear growth compared to the system size increment, eventually becoming disconnected in the thermodynamic limit.

II. NETWORK GENERATION

The basic parameters to define a system of interconnected networks are the subnetworks' size N_i and their average degree $\langle k_i\rangle$. Moreover, the relative size between two subnetworks can be defined as $r_{ij} = N_i/N_j$. Finally, the connection between them, $\langle k_{ij}\rangle$, has to be also under consideration, fulfilling $\langle k_{ji}\rangle = r_{ij}\langle k_{ij}\rangle$.

As we mentioned, from [3], to see multiple percolation transitions the connection between subnetworks must go sublinearly with the system size N^α, with $0 < \alpha < 1$.

Considering three connected networks allows us to play with the structure, designing open and closed chains, not possible in the two networks version. The general scheme is illustrated in FIG. 1.

Now, all that remains is to consider what type of network to use. Focusing on the random graphs for their

simplicity, we would like to highlight the Erdös-Rényi [4] and Gilbert [5] models. They refer to the microcanonical and canonical ensembles respectively.

A. The Erdös-Rényi random graph

The network structure chosen is the Erdös-Rényi [4]. The generation is simple, from a defined $0 \leq p \leq 1$ a link between two vertices is established with probability p and not with $(1-p)$.

To model our subnetworks we only need to define their size N_i and average degree $\langle k_i \rangle$. Then we can calculate the probability $p_i = \langle k_i \rangle / N_i$ to establish the links. Looking over all possible pairs, a real random number uniformly distributed between zero and one is created. If this is lower or equal to the probability just mentioned, p_i, the bond is created. The connections between subnetworks, with $p_{ij} = \langle k_{ij} \rangle / N_j$, are set up following the same steps. Leading to $p_{ij} = p_{ji}$ as it can be easily proved using the subnetworks relations.

The Erdös-Rényi graphs have the same critical exponents as the mean-field model [2], being $\beta = \gamma = 1$ and $\nu = 3$. Furthermore, the percolation threshold of this type of networks, laying down links at random, is at the control parameter value $p_c = 1/\langle k \rangle$.

III. PERCOLATION REALIZATION

In our case, to study the percolation transitions, we have to take the values of the largest cluster G, specifically calculating the average over some realizations $\langle G \rangle$. So the characteristic parameters we need to track are the relative largest component $g \equiv \langle G \rangle / N$ and its fluctuations, this is the susceptibility $\chi \equiv [\langle G^2 \rangle - \langle G \rangle] / \langle G \rangle$. We have taken the alternative susceptibility approach proposed in [3], better for the numerical calculation. Following the mentioned reference, near the critical point, the behaviour of the averaged largest component is

$$\langle G \rangle_c \sim N^{1-\beta/\nu},$$

and, on the other hand, the susceptibility

$$\chi_c \sim N^{\gamma'/\nu},$$

where $\gamma' = \gamma + \beta$. Substituting the critical exponents of the Erdös-Rényi graphs we obtain that $\langle G \rangle_c \sim N^{2/3}$ and $\chi_c \sim N^{2/3}$.

To perform the percolation we applied the Newman-Ziff algorithm, a Monte Carlo simulation described in [6, 7]. It stands out for its applicability to any type of network, being able to perform both site and bond percolation. The algorithm can keep track of the observable quantities of interest in a time that scales linearly with the system size, an improvement over other simulations. Further, it is able to calculate the observables over all the values of the occupation probability, from zero to one.

A. The Newman-Ziff algorithm

First, to set up the bonds uniformly at random, a sorting function is called, establishing a link configuration order. Starting from the beginning of the edges' list, until reaching the last bond M, for each element i another item $i < j \leq M$ is chosen stochastically to exchange positions. Repeating this process is useful to perform different simulations with only one network. Nevertheless, since essentially there is only one sample, the results between simulations will be correlated.

The algorithm used has two main steps to follow. Briefly, from each added edge, we first "find" the clusters involved, those of the two linked vertices. Then, only if the sites belong to different clusters, the "union" is performed. To optimize the process some improvements are made, this is the "weighted union-find with path compression". But first, another important concept is needed. The "tree-based" organization consists of storing the clusters separately, defining a root for each cluster. In this scheme, every vertex has a pointer directly to its root or to other nodes of the group.

Initially, where all the sites are unconnected, every node forms a cluster of size one. This means that every site is a root. Following the configuration list, the bonds are established. In this step, the smart procedure is that the big cluster absorbs the small. Making the root of the small cluster point to the other. This is the "weighted union". To apply it, the size of the clusters must be tracked. For good cluster management, the root labels are set negative, so that they are distinguishable, being equal to minus its cluster size. Conversely, the rest of the cluster's nodes are positive, being their values the sites they point to. On the other hand, if the two groups amalgamated have the same size, the new root is chosen randomly. Last, because traversing a tree to find the root can take a lot of computational effort, all the pointers of the traversed nodes in the "find" statement are set to their root directly. This is the "find with path compression", useful when there are many big clusters to mix, especially near the percolation point.

During the simulation, the observables of interest Q are calculated every time a link is occupied. For this reason, their distribution is discontinuous in the range of study, only for $p = m/M$ values. This corresponds to a microcanonical ensemble, where the specified parameter is the number of established links m, being the energy in statistical physics. If we want to calculate the observable for an arbitrary occupation probability $Q(p)$, the canonical ensemble is needed. With the set $\{Q_m\}$, the values of Q for all possible m, we can calculate

$$Q(p) = \sum_m B(M, m, p) Q_m = \\ = \sum_m \binom{M}{m} p^m (1-p)^{M-m} Q_m, \quad (1)$$

where $B(M, m, p)$ is the binomial distribution.

Meanwhile the program is running, to measure our variables of interest it is necessary to have a separated score for each one.

IV. RESULTS ANALYSIS AND DISCUSSION

We can predict the theoretical behaviour of the simulations. For a network i with a defined $\langle k_i \rangle$, the exact relative largest component for a probability occupation value, taken from [3], is determined by the transcendental equation

$$g_i = 1 - e^{-p\langle k_i \rangle g_i},$$

which can be solved numerically by iterations. Considering the theoretical function $g_i(p)$ of one subnetwork, we can calculate its contribution to the whole system $g'_i(p)$ with

$$g'_i(p) = \frac{N_i}{\sum_l N_l} g_i(p), \tag{2}$$

where l goes through all the subnetworks. Eventually, we can also verify the susceptibility, which describes the phase transitions. Its behaviour through two interconnected networks is given in [3] and is generalized as

$$\chi = \sum_i \frac{\langle G_i \rangle}{\sum_l \langle G_l \rangle} \chi_i. \tag{3}$$

Lastly, it is useful to remember that the subnetwork i percolates at the control parameter value $p_{c,i} = 1/\langle k_i \rangle$.

For a first approach to the problem, we considered the same size for the three subnetworks, $N_1 = N_2 = N_3 = N$. Also, its interconnection, when it is established, is the same. To obtain the desired multiple percolations, the probability of linking two nodes of different networks has to be defined as $p_{ij} = N^{\alpha-2}$, with $0 < \alpha < 1$. When both subnetworks have percolated, as follows from [3], $\langle G_i \rangle, \langle G_j \rangle \sim N$ is fulfilled. At this moment, their connection goes like the wanted, $p_{ij}\langle G_i \rangle \langle G_j \rangle \sim N^\alpha$, presenting the sublinear growth. In short, the structures that we studied, together with the last established parameters and including the average degrees $\langle k_i \rangle$ considered, are represented in FIG. 2. Finally, using the same samples, we averaged 10^4 simulations with different setup orders. We have also applied, to the tracked variables, the equation (1), which considers the binomial distribution of the results.

From FIG. 2, the first structures studied are (a) and (b), very similar to each other. We can calculate the theoretical relative giant component for both cases using (2). This corresponds to the next piecewise function, where the $\langle k_i \rangle$ values are used as labels.

$$g'(p) = \begin{cases} 0 & p < p_{c,50} \\ g'_{50}(p) & p_{c,50} \leq p < p_{c,20} \\ g'_{50}(p) + g'_{20}(p) & p_{c,20} \leq p < p_{c,10} \\ g'_{50}(p) + g'_{20}(p) + g'_{10}(p) & p \geq p_{c,10} \end{cases}$$

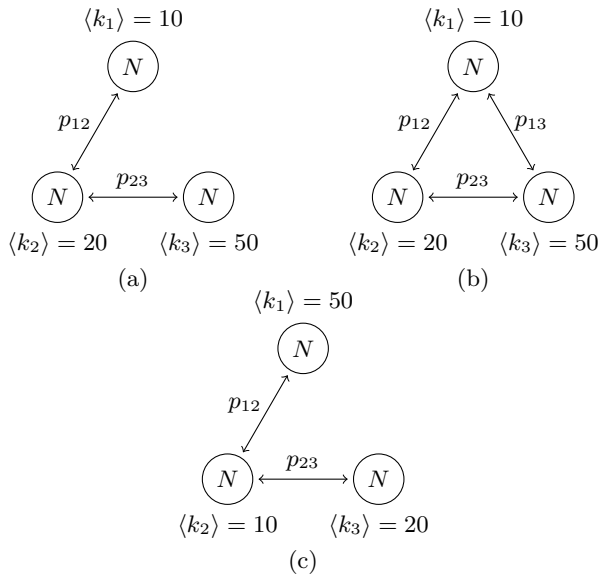

FIG. 2. Structures considered of three networks. The subfigures (a) and (c) are open chains with $p_{12} = p_{23} = N^{\alpha-2}$ and $p_{13} = 0$. Otherwise, (b) forms a closed chain with $p_{12} = p_{13} = p_{23} = N^{\alpha-2}$. The $\alpha = 1/2$ in all the cases.

In these cases, the susceptibility has three peaks. Using (3), the first behaves like the described for the standard percolation transition.

$$\chi_{1\text{st}} \approx \chi_{50} \sim N^{2/3}$$

This is because only the network with $\langle k_3 \rangle = 50$ takes part, going like χ_c. As regards the second peak, corresponding to the percolation of the network with $\langle k_2 \rangle = 20$, the highlighted behaviours are $\chi_{20} \sim \chi_c$ and $\langle G_{20} \rangle \sim \langle G \rangle_c$. Also, because it is sufficiently separated from the first one, we can consider $\chi_{50} \sim$ const. Looking over (3), and since the clustered $\langle k_3 \rangle = 50$ is already extensive, $\langle G_{50} \rangle \sim N$, we obtain the second transition trend.

$$\chi_{2\text{nd}} \approx \chi_{50} + \frac{\langle G_{20} \rangle}{\langle G_{50} \rangle} \chi_{20} \sim \frac{N^{2/3}}{N} N^{2/3} = N^{1/3}$$

Finally, assuming that the third transition point is sufficiently separated from the first and second peaks, we can consider $\chi_{50}, \chi_{20} \sim$ const.

$$\chi_{3\text{rd}} \approx \chi_{50} + \chi_{20} + \frac{\langle G_{10} \rangle}{\langle G_{50} \rangle + \langle G_{20} \rangle} \chi_{10} \sim N^{1/3}$$

We can compare these values with those obtained in FIG. 3 and 4, following the expected behaviour and expanding article [3] results. It is relevant to note that, due to the finite scale, ignoring the contribution of the system with $N = 5000$ to the third peak of the closed chain sample, FIG. 4, we have obtained a better fit.

On the other hand, the following simulation, (c) structure of FIG. 2, corresponds to the connection of two subnetworks through another with a lower average degree.

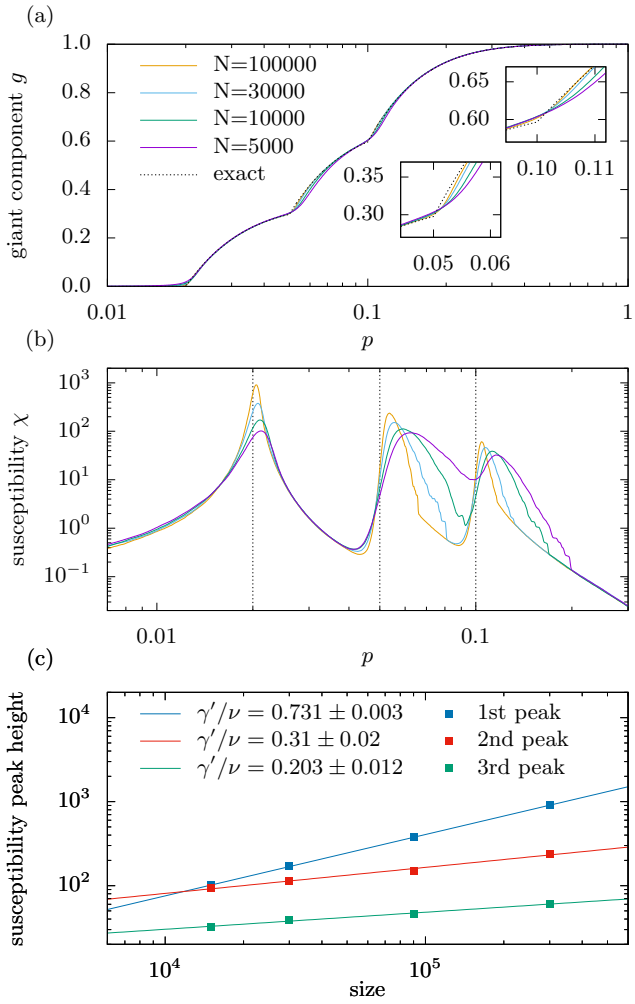

FIG. 3. Results of a bound percolation simulation over a sample of three interconnected networks, with sizes $N_1 = N_2 = N_3 = N$ and average degrees $\langle k_1 \rangle = 10$, $\langle k_2 \rangle = 20$ and $\langle k_3 \rangle = 50$. The modelled structure between subnetworks forms an open chain and it is characterised by the interconnection probabilities $p_{13} = 0$ meanwhile $p_{12} = p_{23} = N^{\alpha-2}$, with $\alpha = 1/2$. Different subnetwork sizes have been simulated. (a) Relative size of the largest connected component g as a function of the bond occupation probability p. The graph includes the numerical exact calculation. (b) Susceptibility χ as a function of the bond occupation probability p, where the theoretical p_c points are highlighted. (c) Susceptibility peak height as a function of the global size. The graph includes the adjusted critical exponents.

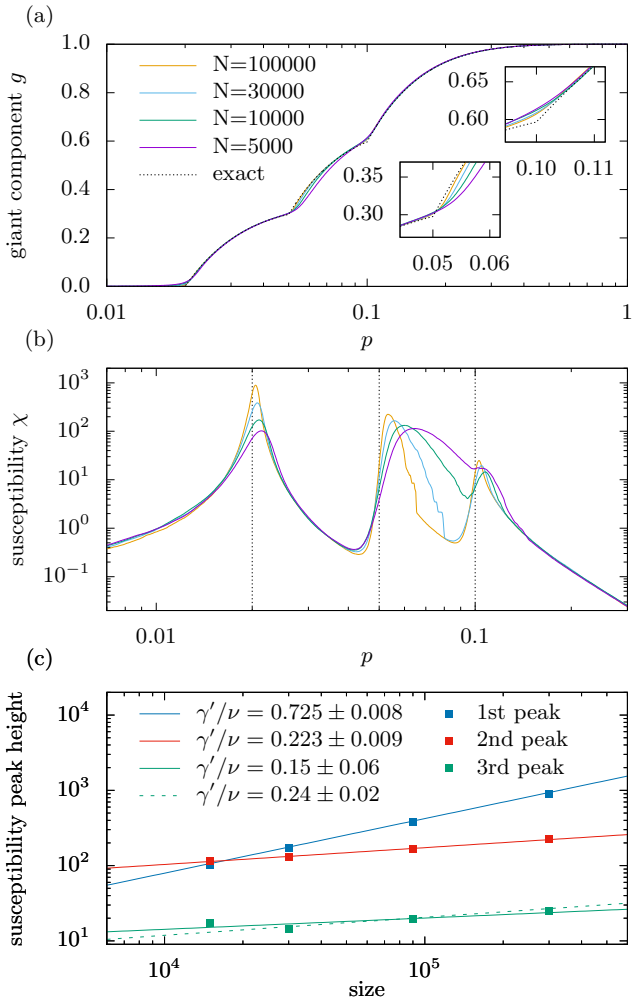

FIG. 4. Results of a bound percolation simulation over a sample of three interconnected networks, with sizes $N_1 = N_2 = N_3 = N$ and average degrees $\langle k_1 \rangle = 10$, $\langle k_2 \rangle = 20$ and $\langle k_3 \rangle = 50$. The modelled structure between subnetworks forms a closed chain and it is characterised by the interconnection probabilities $p_{12} = p_{13} = p_{23} = N^{\alpha-2}$, with $\alpha = 1/2$. Different subnetwork sizes have been simulated. (a) Relative size of the largest connected component g as a function of the bond occupation probability p. The graph includes the numerical exact calculation. (b) Susceptibility χ as a function of the bond occupation probability p, where the theoretical p_c points are highlighted. (c) Susceptibility peak height as a function of the global size. The graph includes the adjusted critical exponents.

Keeping the structure in mind and following the same procedures as seen previously, the theoretical relative giant component can be found.

$$g'(p) = \begin{cases} 0 & p < p_{c,50} \\ g'_{50}(p) & p_{c,50} \leq p < p_{c,20} \\ g'_{50}(p) & p_{c,20} \leq p < p_{c,10} \\ g'_{50}(p) + g'_{20}(p) + g'_{10}(p) & p \geq p_{c,10} \end{cases}$$

This result presents a first suitable difference compared to the previous cases. When the second percolation threshold $p_{c,20}$ is reached, because of the lower degree subnetwork mediation, $\langle k_1 \rangle = 50$ and $\langle k_3 \rangle = 20$ are not merged. Moreover, since g'_{50} is still growing, and therefore greater than g'_{20}, the $\langle k_3 \rangle = 20$ network is not considered. The fusion does not occur until the percolation of the intermediary network.

Taking into account the giant component, the susceptibility presents two peaks in the studied range. The first, as seen, follows the expected critical behaviour.

$$\chi_{1\text{st}} \sim N^{2/3}$$

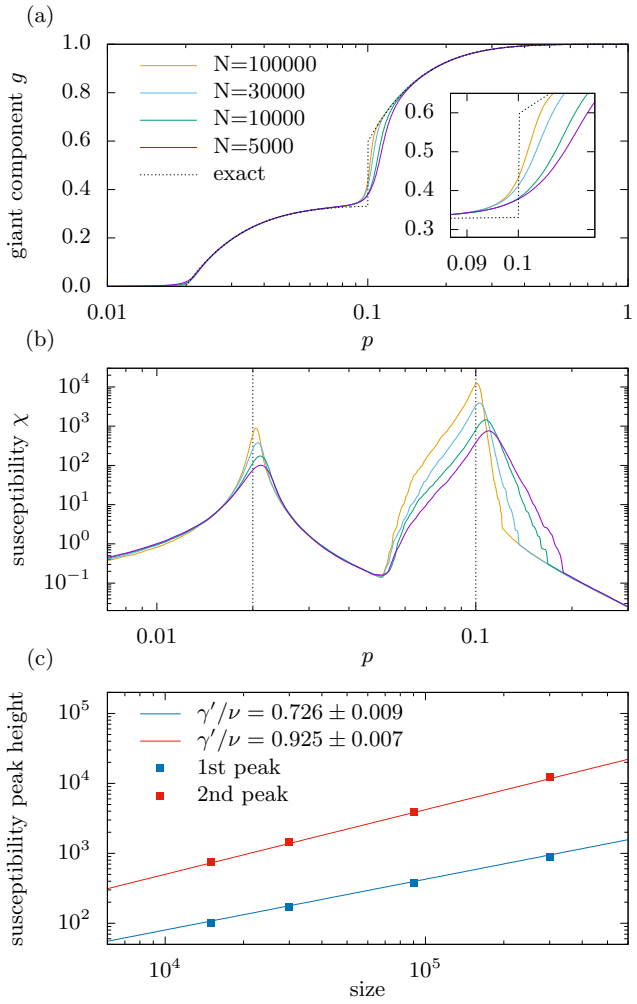

FIG. 5. Results of a bound percolation simulation over a sample of three interconnected networks, with sizes $N_1 = N_2 = N_3 = N$ and average degrees $\langle k_1 \rangle = 50$, $\langle k_2 \rangle = 10$ and $\langle k_3 \rangle = 20$. The modelled structure between subnetworks forms an open chain and it is characterised by the interconnection probabilities $p_{13} = 0$ meanwhile $p_{12} = p_{23} = N^{\alpha-2}$, with $\alpha = 1/2$. Different subnetwork sizes have been simulated. (a) Relative size of the largest connected component g as a function of the bond occupation probability p. The graph includes the numerical exact calculation. (b) Susceptibility χ as a function of the bond occupation probability p, where the theoretical p_c points are highlighted. (c) Susceptibility peak height as a function of the global size. The graph includes the adjusted critical exponents.

Then, although $\langle k_3 \rangle = 20$ percolates, no changes are noticed. It is not until the percolation of the $\langle k_2 \rangle = 10$ subnetwork that a second peak χ_{2nd} appears. Because the subnetwork $\langle k_3 \rangle = 20$ is already clustered, $\langle G_{20} \rangle \sim N$, its suddenly consideration makes $\chi_{\text{2nd}} \sim N$. Leading this union a discontinuous phase transition. The results of this new phenomena are presented in FIG. 5. They also append the experimental critical exponent values, calculated with a linear regression.

V. CONCLUSIONS

With this project, we have proven the existence of multiple percolations in a system of three coupled networks. At the same time, we have found a new discontinuous phase transition as a consequence of the subnetworks structure. Even with the small size of the systems studied compared with the thermodynamic limit, the experimental behaviour is in accordance with the theoretically expected.

Looking towards future studies, it could be interesting to consider different networks' sublinear connections, changing the α value. It also could be possible, in the three subnetworks version, to evaluate different α values at the same time. In fact, introducing more subnetworks would open many possible investigations.

Personally, the research made has been a good approximation to the physics of complex systems. It has allowed me to discover new algorithm logics and delve into some data management techniques.

ACKNOWLEDGMENTS

First, I want to especially express my gratitude towards Marián Boguñá for all his time and effort. For trusting the project, thanks for the explanations, resources and tools provided. Moreover, to introduce me into such an interesting topic.

I appreciate the interest of all my family and friends, for reading the paper and giving me their opinions. Thanks, particularly to my sister, for the final revision. Lastly, I would like to express my gratitude to my parents, to bring me the opportunity of studying this career.

[1] M. E. Newman, *Networks: An Introduction*, (Oxford University Press, Oxford 2010, 1st. ed.).
[2] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, (CRC Press, London 1994, 2nd. ed.).
[3] P. Colomer-de-Simón and M. Boguñá, *Double Percolation Phase Transition in Clustered Complex Networks*, Phys. Rev. X 4, 041020 (2014).
[4] P. Erdös and A. Rényi, *On random graphs*, Publ. Math. Debrecen 6, 290-297 (1959).
[5] E. N. Gilbert, *Random graphs*, Annals of Mathematical Statistics 30 (4), 1141-1144 (1959).
[6] M. E. Newman and R. M. Ziff, *Efficient Monte Carlo Algorithm and High-Precision Results for Percolation*, Phys. Rev. Lett. 85, 4104 (2000).
[7] M. E. Newman and R. M. Ziff, *A fast Monte Carlo algorithm for site or bond percolation*, Phys. Rev. E 64, 016706 (2001).