Controlling the uncertain response of real multiplex networks to random damage

Francesco Coghi, Filippo Radicchi, and Ginestra Bianconi

Controlling the uncertain response of real multiplex networks to random damage

Francesco Coghi,[1] Filippo Radicchi,[2] and Ginestra Bianconi[1]

[1]*School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK*
[2]*Center for Complex Networks and Systems Research,*
School of Informatics, Computing, and Engineering,
Indiana University, Bloomington, IN 47408, USA

We reveal large fluctuations in the response of real multiplex networks to random damage of nodes. These results indicate that the average response to random damage, traditionally considered in mean-field approaches to percolation, is a poor metric of system robustness. We show instead that a large-deviation approach to percolation provides a more accurate characterization of system robustness. We identify an effective percolation threshold at which we observe a clear abrupt transition separating two distinct regimes in which the most likely response to damage is either a functional or a dismantled multiplex network. We leverage our findings to propose a new metric, named safeguard centrality, able to single out the nodes that control the response of the entire multiplex network to random damage. We show that safeguarding the function of top-scoring nodes is sufficient to prevent system collapse.

I. INTRODUCTION

Civil infrastructures, transportation networks, financial networks, as well as cell molecular networks and brain networks, are all good examples of multiplex networks, i.e., complex systems whose topology can be meaningfully represented as a composition of many interacting network layers [1–5]. A central topic in the study of multiplex networks is the characterization of their robustness [6]. This problem is usually approached with percolation theory, where the macroscopic connectedness of the system is studied as a function of the microscopic damage of system elements. The simplest scenario considered in percolation studies of multiplex networks assumes that nodes are initially damaged with probability f (alternatively, one may assumes that nodes are not damaged with probability $p = 1 - f$). Depending on the topology of the system and the value of the probability f, the initial damage of nodes may trigger further damaging avalanches in the system, eventually leading to the complete failure of the multiplex [6]. On networks with infinite size, it has been shown that percolation yields a discontinuous hybrid transition of the Mutually Connected Giant Component (MCGC), thus radically different from the usual continuous transition observed in isolated networks [6–21]. The discontinuity of the transition indicates that multiplex networks are significantly more fragile than their single layers taken in isolation. The reason is that, at the percolation transition, a multiplex network is affected by large avalanches of failure that suddenly dismantle the whole network leading to a discontinuous phase transition [6]. This result is central in the study of percolation and is playing a major role in the active research field aiming at identifying dynamical rules that can change the nature of phase transitions from continuous to discontinuous [22–27].

Percolation theory, on single-layer as well as on multiplex networks, is traditionally studied in the *mean-field* approach by characterizing the average response of a network to initial damage [28, 29]. This approach is totally justified in the infinite network limit where percolation is self-averaging, i.e., the fluctuations around the mean behaviour are vanishing. However, the interest in the percolation transition is often driven by applications which always involve finite (and sometime not too large) networks [6, 14, 19, 30]. Further in practical applications, the prediction of eventual, even if extremely rare, catastrophic failures is way more important than the characterization of the average behavior of a system.

To provide a pragmatic characterization of the response to damage of real networks, recent papers, such as Refs. [31–34], on percolation in single-layer networks went beyond the standard *mean-field* approach. In Ref. [32], a theoretical framework based on large deviation theory was proposed to predict the probability distribution $\pi(R)$ for the relative size R of the giant component in single instances of the percolation model on a given network. The approach allows for the theoretical computation of $\pi(R)$ starting from any real network datasets. Results of the paper show that the average value of $\pi(R)$, hence the main observable of the mean-field approach, may not be the best metric to study system robustness. Also, optimal percolation defines a problem that goes beyond the traditional percolation model [33, 34]. Optimal percolation refers to the identification of the *optimal (minimal) structural node set* whose removal leads a destruction of the entire network. In this sense, optimal percolation is the problem of identifying the one rare realization of damage that has the most dramatic consequences for the network.

In the context of multiplex networks several work have started to characterize the response to damage beyond the mean-field approach. In Ref. [35], the authors studied finite-size effects in multiplex network percolation focusing both on the final size of the MCGC and the avalanche distribution. In Ref. [36], Kitsak and collaborators analyzed the stability of the MCGCs by considering the overlap among a large number of MCGCs resulting from initial damage configurations drawn from the same distribution. Optimal percolation was recently

extended to multiplex networks in Ref. [37]. The main finding is that optimal percolation on a multiplex network is a rather distinct problem from the one defined for the individual network layers that compose the multiplex. Further work in this direction was presented in Ref. [38].

In this paper, we aim at providing a novel characterization of the percolation transition in multiplex networks. The approach we propose is similar to one already used in Ref. [32] for isolated networks, thus placing emphasis on large deviation properties of percolation. We consider a process where a fraction $f = 1 - p$ of nodes is initially damaged, and show that the probability $\pi(R)$ that the relative size of the MCGC equals R is bimodal. The two peaks of the distribution $\pi(R)$ correspond to the percolating and non-percolating phases, and in specific ranges for the parameter p they quantify an equal likelihood for the system to be in the functioning or nonfunctioning regimes. In this respect, the mean-field percolation diagram where the average value $\bar{R}$ is plotted against p provides distorted information about the robustness of the system, making it look less fragile than actually is. An alternative phase diagram can be instead created by replacing $\bar{R}$ with $\hat{R}$, i.e., the mode of $\pi(R)$. In the phase diagram, $\hat{R}$ displays a clear discontinuity. However, the large fluctuations observed in single instance of percolation cannot be captured entirely by this single metric. The actual robustness properties of the network can be fully understood only by looking at the distribution $\pi(R)$. Nevertheless, as a matter of fact, the combination of the two metrics $\bar{R}$ with $\hat{R}$ results in an effective tool for the characterization of its robustness profile. These numbers respectively underestimate and overestimate the fragility of the system, allowing for a direct interpretation the fragility properties of the network in practical contexts. In this respect, we identify an *effective critical point* p_c with the discontinuity of $\hat{R}$, and show that, for $p = p_c$, the system is characterized by significant uncertainty on the possible outcomes of the percolation process. Further, we propose a score, named safeguard centrality (SC), to identify the nodes that have major influence in safeguarding the MCGC at criticality. We find that the set of top nodes according to SC has a very significant overlap with the sets identified as solutions to the optimal percolation problem.

ant component of the multiplex network formed by the largest set of nodes in which each pair of nodes is connected by at least a path in each layer of the multiplex networks (where all these paths must remain inside the MCGC) [6, 7].

In an infinite multiplex network, the MCGC is an extensive component in the sense that it includes a non-vanishing fraction of all the nodes. The same exact definition doesn't apply to finite real networks. However, it is usual practice in the field to use the expression MCGC in a finite multiplex network to indicate its largest mutually connected component. We will interpret this component as the *giant* one only when its size exceeds $\sqrt{N}$.

To study the robustness of a multiplex network, we employ a generalized percolation model where nodes are initially damaged with probability $f = 1 - p$ and the relative size R of the MGCC is monitored as a function of p [6]. The characterization of the robustness of the multiplex network thus reduces to the study of the generalized percolation transition. On infinite networks, the transition is investigated by studying the average fraction $\bar{R}$ of nodes in the MCGC as a function of p. This critical phenomenon displays noticeable properties [6, 7]. The MCGC emerges with a discontinuous hybrid phase transition at $p = p_c$ where the multiplex network is affected by avalanches of failures propagating back and forth among the different layers. This transition has been fully characterized on multiplex networks with Poisson and scale-free degree distributions without edge overlap [6, 7]. In particular, the transition is always discontinuous and hybrid, and interdependent multiplex networks are significantly more fragile than their single layers taken in isolation [1, 6, 7]. Recently, it has been shown also that multiplex network models with edge overlap although they tend to be somewhat more robust than multiplex networks without overlap, they present always discontinuous hybrid phase transitions [16, 17, 20].

The percolation model applied to multiplex networks is particularly relevant in robustness studies of real interdependent multiplex networks [15, 19]. However, in a large variety of cases, multiplex networks are far from the large network limit. It is therefore essential to understand whether the average fraction $\bar{R}$ of nodes in the MCGC is a suitable metric to assess the robustness of real interdependent multiplex networks.

II. PERCOLATION OF INTERDEPENDENT MULTIPLEX NETWORKS

We consider a multiplex network $\vec{G} = (G^{[1]}, G^{[2]})$ formed by $M = 2$ layers and N nodes [2, 5]. Each layer $\alpha = 1, 2$ consists of a network $G^{[\alpha]} = (V, E^{[\alpha]})$. The set V of N nodes is identical for both layers. The set of links $E^{[\alpha]}$ is instead typical of the layer α. We monitor the connectedness of the interdependent multiplex network by looking at the size of the Mutually Connected Giant Component (MCGC) [6]. The MCGC is the gi-

III. WHY A LARGE DEVIATION APPROACH TO PERCOLATION IS NEEDED

Percolation theory should provide us with a good prediction of the behavior of the size of the MCGC as a function of the probability $f = 1 - p$ of damaging nodes at random. The prediction should include a good estimate of the percolation threshold p_c. Knowing the value of this quantity is in fact very informative as it enables us to judge whether the considered multiplex network displays a nonvanishing MCGC or not. Also, the predic-

tion should inform us about the nature of the percolation transition, whether this is smooth or abrupt. The simple analysis conducted below reveals that the mean-field theory of percolation, when applied to real multiplex datasets, may fail to provide us with reliable information, as the average fraction of nodes $\bar{R}$ in the MCGC is a metric that does not satisfy the desired requirements of an informative percolation theory.

In analysis, we considered several real-world multiplex networks, including air transportation networks among major US air carriers (Delta Airlines-United Airlines; American Airlines-United Airlines and United Airlines-Delta Airlines) [19], and biological networks (the genomic network of the D. Melanogaster [40] and the C. Elegans connectome [39, 40]). Basic properties of these multiplex networks are reported in Table I. Number of nodes N range between 73 and 557, thus showing that multiplex networks of practical interest may be small/medium sized systems. Further, the comparison between number of links $L^{[\alpha]}$ in each layer α and total number of multilinks [5] $L^{(1,0)}, L^{(0,1)}, L^{(1,1)}$, indicating the pair of nodes connected only in layer 1 ($L^{(1,0)}$) only in layer 2 ($L^{(0,1)}$) or in both layers ($L^{(1,1)}$), emphasizes that the level of link overlap in real multiplex networks may vary from system to system. We simulated a single realization of the random percolation model by damaging nodes sequentially with increasing probability $f = 1-p$. In a single realization, we assign a random uniformly distributed variable x_i to each node i of the multiplex networks, and, for each value of f, we damage all nodes i such that $x_i \leq f$. A single realization of percolation can be described by the dependence of the relative size R of the MCGC as a function of p.

In Figure 1, we provide evidence of the large fluctuations observed for single realizations of percolation. The inaccuracy of the mean-field approach to percolation in predicting the robustness of real multiplex networks of small/medium size is apparent. For each of the considered datasets, we compare the average size $\bar{R}$ of the MCGC and the size R of the MCGC of two single realizations of percolation. As the figure shows, $\bar{R}$ performs poorly with respect to the informative requirements that we mentioned above. First, the size R of the MCGC for a given value of p has strong fluctuations around the mean $\bar{R}$. Second, the position of the percolation threshold p_c inferred from $\bar{R}$ can be dramatically misleading, providing only a lower bound to the actual transition points observed in single realizations of the model. This fact arises because, for any given p, $\bar{R}$ is positive also when in most of the realizations the size of the MCGC is zero, being $\bar{R}$ the average value of non-negative numbers. Finally, the abrupt nature of the transitions associated with individual realizzations is not well captured by $\bar{R}$, which instead displays a continuous behavior.

In summary, the mean-field observable $\bar{R}$ underestimates the true fragility of a real multiplex network. Correct predictions can be achieved only with an approach that actually accounts for large deviations.

IV. LARGE DEVIATION APPROACH TO PERCOLATION

In this paragraph, we establish the general theoretical framework for characterizing the large deviation properties of percolation in interdependent multiplex networks. Our goal is to quantify the response of a multiplex network to an initial damage of the nodes using a metric different from the mere average fraction $\bar{R}$ of nodes in the MCGC. In particular, we will explore the properties of the entire distribution $\pi(R)$ of observing a MCGC formed by a fraction R of nodes. The distribution $\pi(R)$ will be studied as a function of p, i.e., the probability that a node is not initially damaged.

We consider a large number Q of random initial damage realizations. Each initial damage configuration $\mu = 1, 2, \ldots, Q$ is denoted by $\{s_i^\mu\}_{i=1,2,\ldots,N}$, where $s_i^\mu = 0$ if node i is initially damaged, and $s_i^\mu = 1$, otherwise. We assume that each node is damaged independently with probability $f = 1 - p$. Therefore, the probability associated to the initial damage realization $\{s_i^\mu\}$ is

$$\mathcal{P}(\{s_i^\mu\}) = \prod_{i=1}^N \left[p s_i^\mu + (1-p)(1-s_i^\mu) \right]. \tag{1}$$

For each initial damage configuration, we determine whether node i belongs to the MCGC, i.e., $\sigma_i^\mu = 1$, or not, i.e., $\sigma_i^\mu = 0$. The fraction R^μ of nodes in the MCGC is

$$R^\mu = \frac{1}{N} \sum_{i=1}^N \sigma_i^\mu. \tag{2}$$

For any given value of p, different initial damage configurations might induce MCGCs of different sizes. In order to study the distribution $\pi(R)$ of the fraction of the nodes in the MCGC for a random realization of the initial damage $\{s_i^\mu\}$ with probability $\mathcal{P}(\{s_i^\mu\})$ we consider a large number P of realizations of the initial damage and we estimate $\pi(R)$ as

$$\pi(R) = \frac{1}{Q(\Delta R)} \sum_{\mu=1}^Q \delta(R, R^\mu), \tag{3}$$

where $\delta(x,y) = 1$ if $x = y$, and $\delta(x,y) = 0$, otherwise. $\Delta R = 1/N$ represents the width of the bins used for estimating the distribution, thus providing the proper normalization condition for $\pi(R)$. From the full distribution $\pi(R)$ of the sizes R of the MCGCs, it is possible to extract two major statistical quantities: the average size of the MCGC, namely $\bar{R}$, and the typical (most probable) size of the MCGC, namely $\hat{R}$. The quantities are defined respectively as

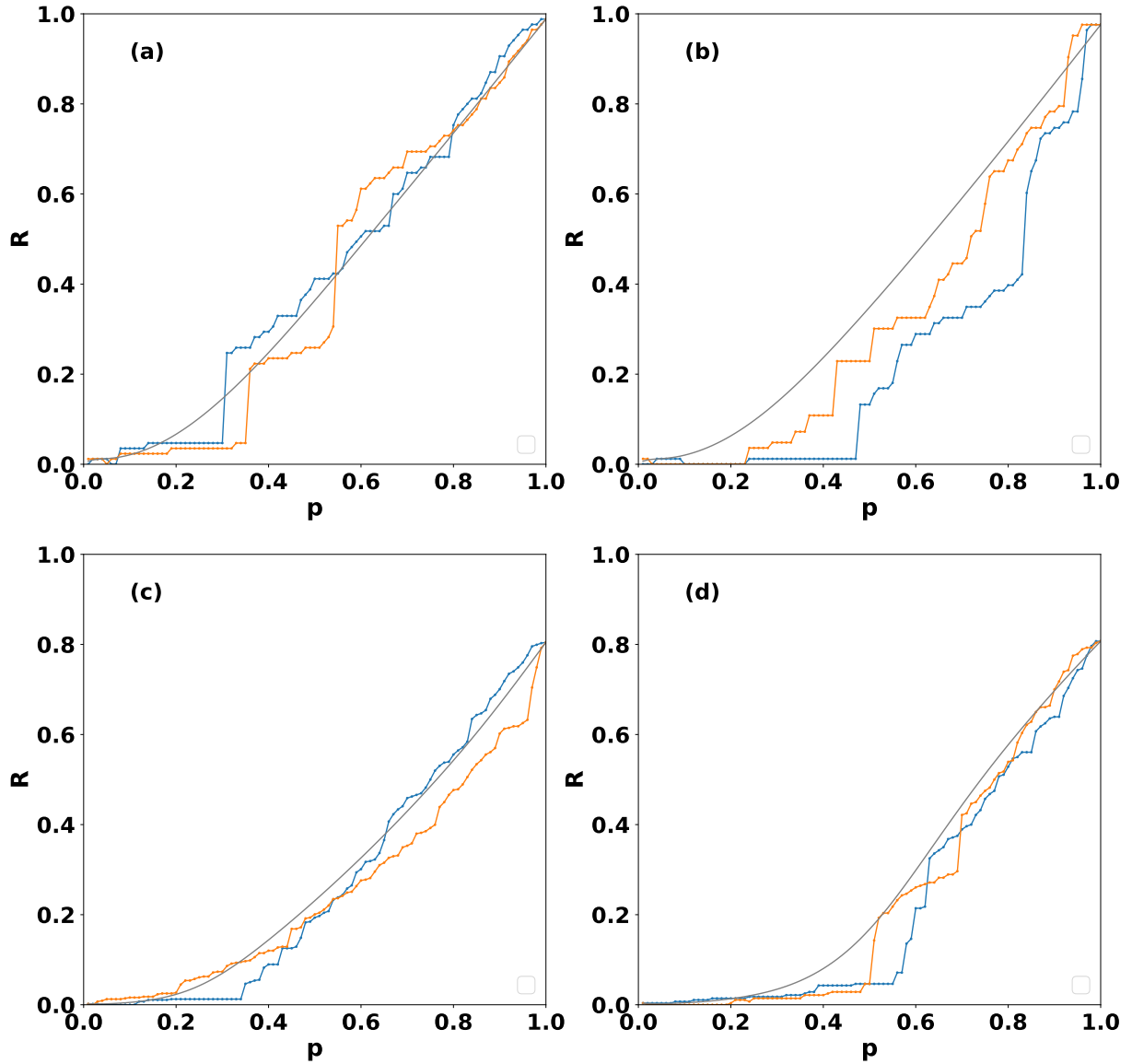$$\bar{R} = \frac{1}{N} \sum_R R\, \pi(R) \tag{4}$$

FIG. 1: **The mean-field approach to percolation can be misleading for characterizing the robustness of finite multiplex networks.** Single realizations of the percolation process described by the corresponding size R of the MCGC as a function of p (blue and orange curves) are plotted together with the average size $\bar{R}$ or the MCGC measured over 10^6 realizations of the initial damage. The different panels correspond to four different datasets: the American Airlines-Delta Airlines multiplex network (panel a), Delta Airlines-United Airlines multiplex network (panel b), the Drosophila Melanogaster genetic network (panel c), and the C. Elegans connectome (panel d).

and

$$\hat{R} \;=\; \arg\max_R \,[\pi(R)]. \tag{5}$$

We stress once more that all quantities defined above are defined given the probability $f = 1-p$ for the random initial damage of each node. We avoid to write explicitly such a dependence just for shortness of notation.

V. LARGE DEVIATION OF PERCOLATION IN REAL MULTIPLEX NETWORKS

A. Typical versus average size of the MCGC

In order to explore the large deviation properties of percolation on real multiplex network we considered again the real-world multiplex networks listed in Table I. We have calculated $\pi(R), \bar{R}, \hat{R}$ by performing numerically $Q = 10^6$ realizations of the initial damage as a function of p. Our results reveal that the typical response to

Duplex	N	$L^{[1]}$	$L^{[2]}$	$L^{(1,0)}$	$L^{(0,1)}$	$L^{(1,1)}$
American-United airlines	73	229	270	161	202	68
American-Delta airlines	84	258	442	190	374	68
United-Delta airlines	82	282	404	226	348	56
D. Melanogaster genomic network	557	1421	1164	953	696	468
C. Elegans connectome	279	514	888	403	777	111

TABLE I: **Main properties of the studied datasets.** For each analysed dataset we indicate: the total number of nodes N, the total number of links $L^{[1]}$ in layer 1, the total number of links $L^{[2]}$ in layer 2 and the total number of multilinks $L^{(1,0)}, L^{(0,1)}, L^{(1,1)}$ indicating the number of pairs of nodes connected only in layer 1, only in layer 2 or in both layers, respectively.

damage $\hat{R}$ uncovers a completely different scenario with respect to the one indicated by the average response to damage $\bar{R}$ for each of the studied datasets (see Figure 2). Indeed, while $\bar{R}$ decreases smoothly for decreasing values of p, suggesting that the system might be robust to damage, $\hat{R}$ reveals a discontinuous behaviour with a rapid jump of $\hat{R}$ from $R = R_c \gg 1/N$ to $R = 1/N$ at the *effective critical threshold* $p = p_c$ where $\pi(1/N) = \pi(R_c)$. Hence, $\hat{R}$ highlights a risk of systemic failure that is not visible from $\bar{R}$, pointing out a serious shortcoming of the average metric in characterizing the true fragility of the system.

B. Bimodality of the distribution of the size of the MCGC

In order to fully characterize the large deviation properties of percolation, we need to study the probability distribution $\pi(R)$ of the size R of the MCGC. The characterization of this distribution as a function of the probability p will reveal why $\hat{R}$ has a discontinuous behavior as a function of p. Here we have considered in detail the American Airlines-United Airlines dataset. For this dataset, the probability distribution $\pi(R)$ can be studied together with $\bar{R}$ and $\hat{R}$ as a function of p (see Figure 3). Starting from high values of p and decreasing p, we observe that initially the distribution $\pi(R)$ is unimodal, and the most likely outcome $\hat{R}$ decreases. However, for lower values of p, the distribution $\pi(R)$ becomes bimodal and for $p = p_c \simeq 0.40$ it has two maxima at $R = R_c \simeq 0.27$ and $R = 1/N \simeq 0.014$ with $\pi(1/N) = \pi(R_c)$. Finally, for even lower values of p, i.e., for $p < p_c$, $R = 1/N$ becomes the most likely size of the MCGC. We checked that the emergence of a bimodal distribution is not an artifact of specific correlations build in the multiplex network structure. In fact, we tested that the feature of the distribution is unaffected by different randomization procedures of the structure of the multiplex (see Supplementary Information [41]).

We stress that the large deviation of percolation consists in the full characterization of the distribution $\pi(R)$, and not exclusively on the characterization of the typi-

cal size $\hat{R}$ of the MCGC. If we just based our prediction about the robustness of a real multiplex network on the observation of the typical size $\hat{R}$ of the MCGC, we would overestimate its fragility. In fact, even for $p < p_c$ so that $\hat{R} = 0$, there is a significant probability that the MCGC is non-vanishing. In this regime, $\pi(R)$ is still bimodal, and the width of the distribution around the two peaks is not symmetric. As such, even if the left peak is higher than the right one, still the right mode of $\pi(R)$ may have higher weight. The relative weight of these two modes is clearly an important information for assessing the robustness of the multiplex network as a whole. In Figure 4, we propose an analysis that accounts for them. Indicate with R_{min} the position of the local minimum in the distribution $\pi(R)$ separating the two peaks. Then evaluate the relative weight of the two modes by measuring the probability $P(R < R_{min})$ and its complementary probability $P(R \geq R_{min})$. The metrics reveal that at $p = p_c$ the probability $P(R \geq R_{min})$ is still larger than 50%, thus providing information that is not directly retrievable either from $\hat{R}$ or $\bar{R}$. In principle, any scalar observable extracted from the distribution $\pi(R)$ can provide new information about the system robustness. However, we argue that the combined use of $\hat{R}$ and $\bar{R}$ generates already sufficient knowledge for many practical tasks. These metrics have in fact the nice feature of representing respectively an over-estimation and under-estimation of the fragility of the multiplex network, thus identifying the range of p values where the percolation transition is expected to happen. This interpretation of $\hat{R}$ and $\bar{R}$ will become more precise in Sec VI where we address finite-size effects, and we will show that in the limit of infinite networks the two metrics become identical.

C. Susceptibility, Specific heat and Correlations

To further characterize the properties of the system, we studied correlations existing among the states of different nodes. This study is interesting from two different points of view. First, correlations are not captured by the mean-field approach to percolation; hence, their analysis allows us to obtain information that is usually
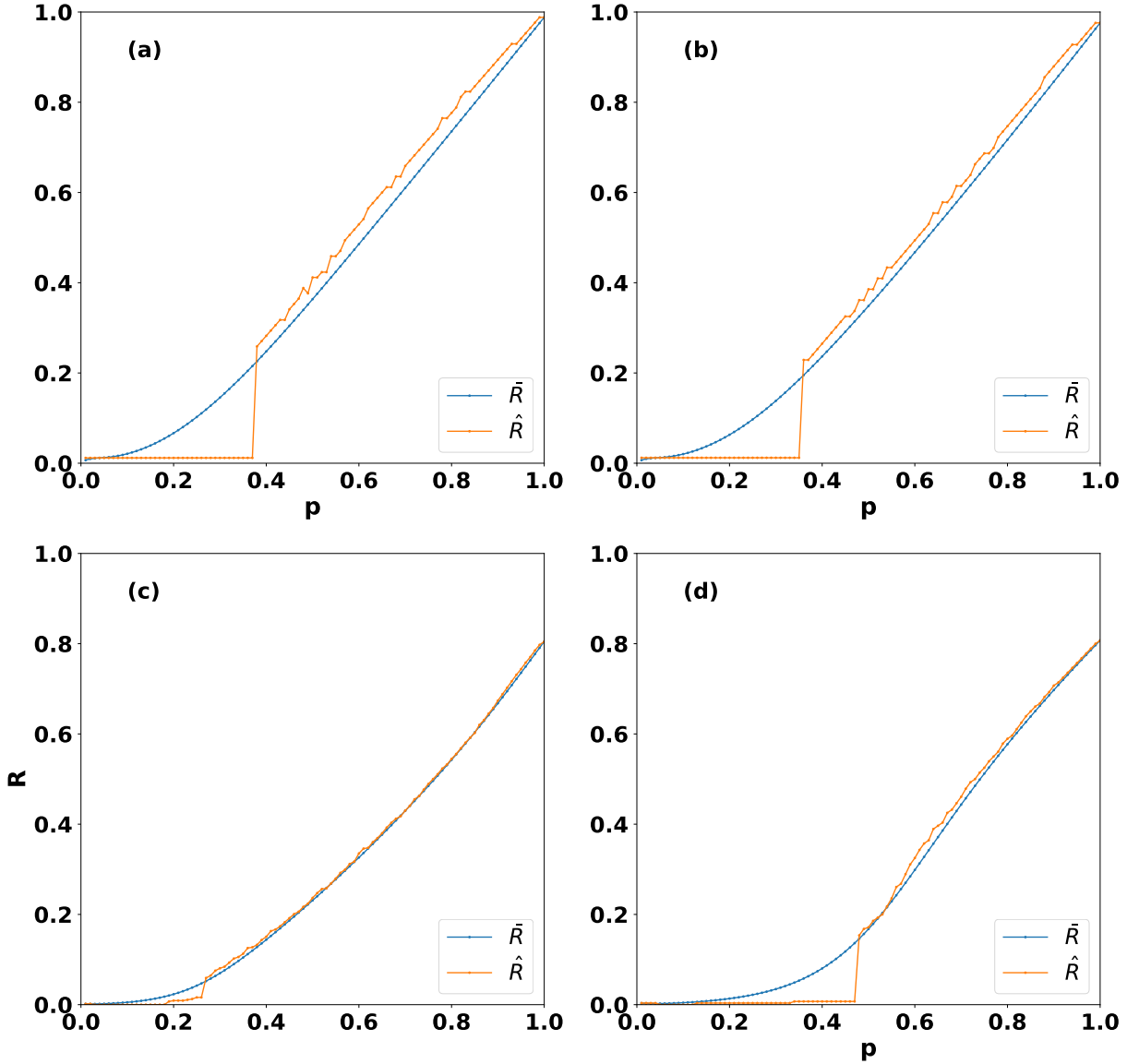
FIG. 2: **Typical versus average size of the MCGC.** We report $\bar{R}$ and $\hat{R}$ for four different data sets: the American Airlines-Delta Airlines multiplex network (panel a) Delta Airlines-United Airlines multiplex network (panel b), the Drosophila Melanogaster genetic network (panel c) the C. Elegans connectome (panel d). The curves $\bar{R}$ vs. p and $\hat{R}$ vs. p have been numerically calculated from the distributions $\pi(R)$ of observing a MCGC with relative size R. The distribution $\pi(R)$ is constructed for each value of p by performing P initial realizations of the damage. We use $Q = 10^6$ for all the multiplex network datasets.

neglected in standard analyses. Second, average correlations between nodes allows us to define a well-grounded version of the susceptibility for percolation on multiplex networks. This definition is in line with the usual practice adopted in the study of critical phenomena, where the susceptibility of a system is taken proportional to the average correlation.

We define the average correlation χ between the state of every pair of nodes in different realizations of the damage and the average correlations χ_{NN} calculated only among of neighboring nodes in at least one layer of the multiplex network. These quantities are defined respectively as

$$\chi = \frac{1}{N(N-1)} \sum_{i \neq j} \left[\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \right],$$

$$\chi_{NN} = \frac{1}{\langle k \rangle N} \sum_{i=1}^{N} \left[\sum_{j \in \mathcal{N}_i} \left[\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \right] \right], \quad (6)$$
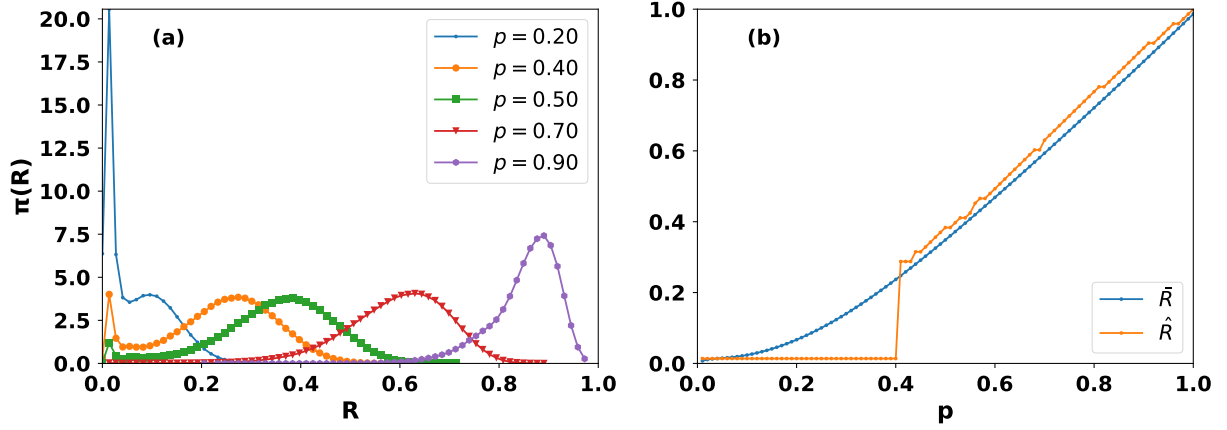
FIG. 3: **The large deviation study of percolation on the American Airlines-United Airlines duplex network.** The probability distribution $\pi(R)$ of observing a MCGC of relative size R in the American Airlines-United Airlines duplex network is shown in panel (a) for different values of p. The average $\bar{R}$ and the most likely $\hat{R}$ size of the MCGC of the same dataset are plotted as a function of p in panel (b). These results are obtained from numerical simulations of $Q = 10^6$ random initial realizations of the damage.
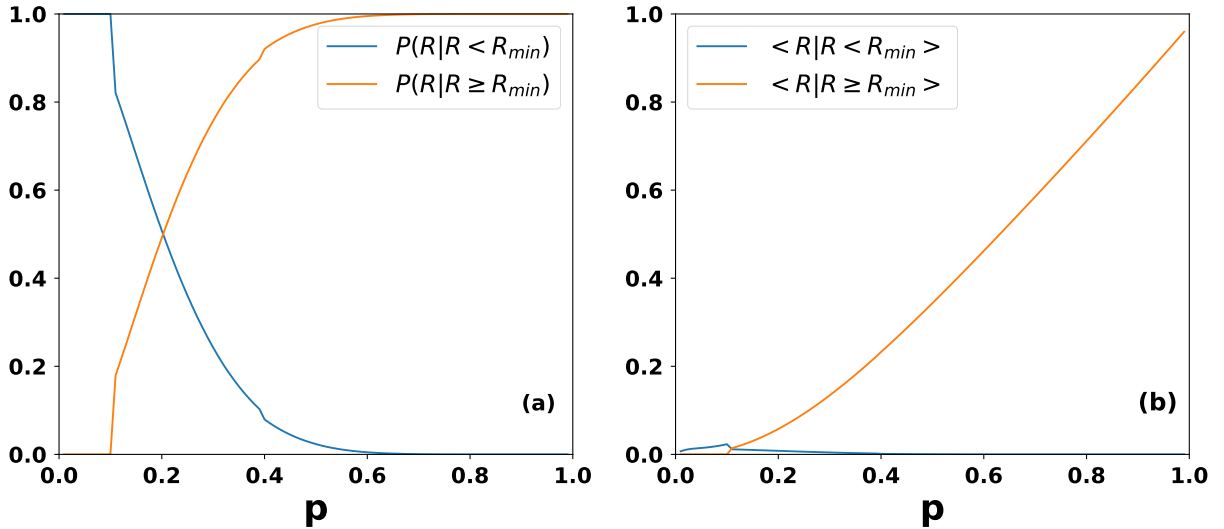

FIG. 4: **The characterization of the two modes of the distribution $\pi(R)$.** The probabilities $P(R \geq R_{min})$, $P(R < R_{min})$ (panel a) and the average size $\langle R|R \geq R_{min}\rangle$ and $\langle R|R > R_{min}\rangle$ of the MCGC corresponding to the two modes (panel b) are plotted as a function of p for the American Airlines-United Airlines duplex network. These results are obtained by performing $Q = 10^6$ realizations of the initial damage.

where we indicated with $\langle\sigma_i\rangle$ and $\langle\sigma_i\sigma_j\rangle$ the averages

$$\langle\sigma_i\rangle = \frac{1}{Q}\sum_{\mu=1}^{Q}\sigma_i^\mu,$$

$$\langle\sigma_i\sigma_j\rangle = \frac{1}{Q}\sum_{\mu=1}^{Q}\sigma_i^\mu\sigma_j^\mu, \qquad (7)$$

and we indicated with $\mathcal{N}_i$ the set composed by all neighbors of node i in at least one layer of the multiplex network. Moreover, we have evaluated the recently introduced *specific heat* C of percolation [31] with $C = Nc$

and c defined as

$$c = \frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle(1-\langle\sigma_i\rangle) \qquad (8)$$

indicating the average fluctuations of the state of a single node. The specific heat C together with the correlation coefficient χ determines the variance σ_R^2 of the size of the
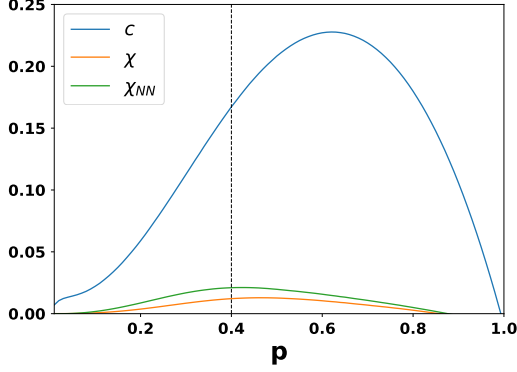
FIG. 5: **Correlations and specific heat for the American Airlines-United Airlines duplex network.** The specific heat of percolation C, the correlation coefficient χ_{NN} among neighbor nodes, and the correlation coefficient χ among any pair of nodes calculated for the American Airlines-United Airlines dataset are plotted as a function of p. The solid dashed line indicates the effective critical point $p = p_c$. These results are obtained from numerical simulations of $Q = 10^6$ random realizations of the initial damage.

giant component R. In fact we have

$$
\begin{aligned}
\sigma_R^2 &= \frac{1}{N^2} \sum_{i,j} \left[\langle \sigma_i \sigma_j \rangle - \sum_{i,j} \langle \sigma_i \rangle \langle \sigma_j \rangle \right] \\
&= \chi \left(1 - \frac{1}{N} \right) + \frac{C}{N^2}.
\end{aligned}
\tag{9}
$$

In Figure 5, we plot c, χ and χ_{NN} as a function of p for the American Airlines-United Airlines duplex network. From this figure, it is possible to show that these curves display a maximum as a function of p. We note that for very small values of p, when the MCGC is also very small, the correlation coefficients and the specific heat are expected to be small since typically most of the nodes will be damaged. Similarly when p is approaching one, most of the nodes will be undamaged yielding small correlations and specific heat. Therefore, the observed maximum of c, χ and χ_{NN} as a function of p is expected.

For a good susceptibility measure, we would like to observe a maximum in correspondence of the percolation threshold. We note that the maximum of χ_{NN} is achieved for values of p that are closer to the transition point $p = p_c$ than those corresponding to the maximum of χ. Additionally from Figure 5, we can notice that correlations among nearest neighbors are, on average, higher than the correlations among any pair of nodes, i.e. $\chi_{NN} \geq \chi$.

Overlap between MCGCs

We have emphasized that, on finite networks, MCGCs resulting from two initial damage configurations drawn from the same distribution $\mathcal{P}(\{s_i^\mu\})$ can have different relative size R. In order to quantitatively evaluate how similar are two different MCGCs resulting from two different configurations μ and ν of the initial damage we propose to use the *overlap* $q^{\mu,\nu}$. The overlap $q^{\mu,\nu}$ is given by the sum between fraction of nodes that belong to both MCGCs and the sum of nodes that do not belong to the MCGC for both realizations μ and ν of the initial damage, i.e.,

$$
q^{\mu,\nu} = \frac{1}{N} \sum_{i=1}^{N} \left[\sigma_i^\mu \sigma_i^\nu + (1 - \sigma_i^\mu)(1 - \sigma_i^\nu) \right] ,
\tag{10}
$$

where $\sigma_i^{\tilde\mu} = 1$ ($\sigma_i^{\tilde\mu} = 0$) indicates that node i is in (is not in) the MCGC after the initial damage configuration $\tilde\mu$ with $\tilde\mu \in \{\mu, \nu\}$. By definition, we have that $q^{\mu,\nu} \in [0,1]$. Values of overlap close to one indicate that the two MCGCs have a very similar node composition, while values of the overlap close to zero indicate that the two MCGCs are very different in terms of node composition. For values of p close to one, where most of the nodes belong to the MCGC, and for values of p close to zero, where most of the nodes do not belong to the MCGC, the typical overlap among MCGCs is expected to be high; instead, typical overlap values should be small for intermediate values of p. We evaluated the average value $\bar{q}$ and the standard deviation $\sigma_{\bar{q}}$ of the overlap measured for different values of p. These metrics are computed using $\tilde{Q}$ pairs of realizations of the initial damage (μ_n, ν_n) (with $0 < n \leq \tilde{Q}$), and using the definitions

$$
\begin{aligned}
\bar{q} &= \frac{1}{\tilde{Q}} \sum_{n=1}^{\tilde{Q}} q^{\mu_n, \nu_n}, \\
\sigma_{\bar{q}} &= \frac{1}{\tilde{Q}} \sum_{n=1}^{\tilde{Q}} \left(q^{\mu_n, \nu_n} - \bar{q} \right)^2 .
\end{aligned}
\tag{11}
$$

We note that $\bar{q}$ is expected to be strongly correlated with the specific heat $C = Nc$ [31]. In fact, for large values of $\tilde{Q}$, we can approximate $\bar{q}$ with

$$
\begin{aligned}
\bar{q} &\simeq \frac{1}{N} \sum_{i=1}^{N} \left[\langle \sigma_i \rangle^2 + (1 - \langle \sigma_i \rangle)^2 \right] \\
&= 1 - 2c.
\end{aligned}
\tag{12}
$$

In Figure 6, we report $\bar{q}$ and $\sigma_{\bar{q}}$ calculated over $\tilde{Q} = 10^6$ pairs of random realizations of the initial damage performed over the American Airlines-United Airlines multiplex networks. We observe that Eq. (12) is satisfied. Therefore, $\bar{q}$ has a minimum corresponding to the maximum of c. Interestingly, $\sigma_{\bar{q}}$ displays two maxima as a function of p, and achieves its absolute maximum for values of p preceding $p = p_c = 0.4$, i.e., $p = 0.34$.

Nevertheless, the full distribution of the overlap $\rho(q)$ observed over $\tilde{Q}$ pairs of realizations of the initial damage encodes more information than its average $\bar{q}$. In particular, the distribution reflects the many-body correlations
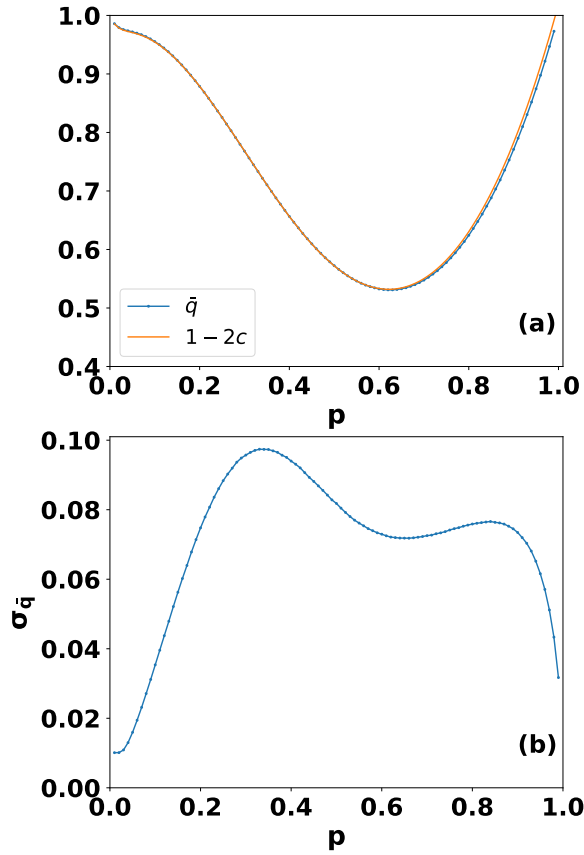
FIG. 6: The mean value $\bar{q}$ (panel a) and the standard deviation $\sigma_{\bar{q}}$ (panel b) of the overlap distribution $\rho(q)$ are plotted versus p for the American Airlines-United Airlines multiplex network. These statistical properties have been numerically evaluated starting from $\tilde{Q} = 10^6$ pairs of random realizations of the initial damage.
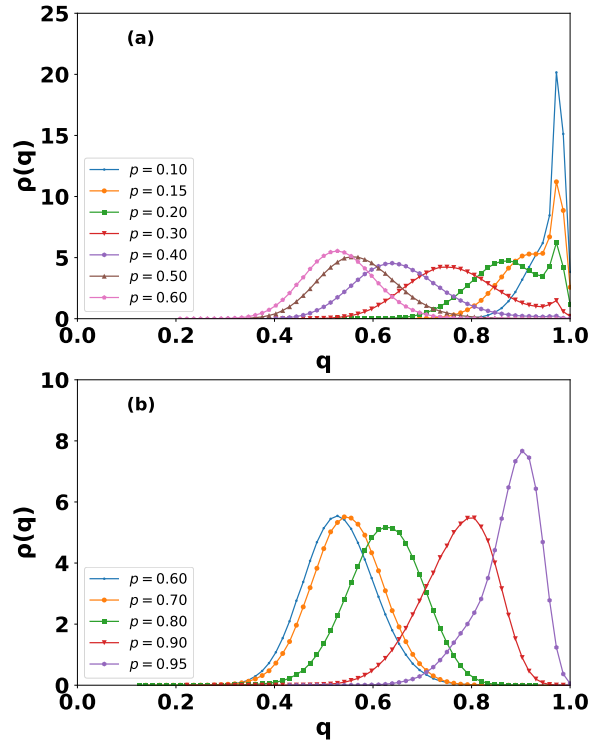

FIG. 7: The overlap distributions $\rho(q)$ among the MCGCs of the American Airlines-United Airlines duplex network are plotted for values of p given by $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$ (panel a) and $p = 0.6, 0.7, 0.7, 0.9, 0.99$ (panel b). These results indicate that for $p < 0.6$ the typical overlap among MCGCs decreases with increasing values of p while for $p > 0.6$ the typical overlap among MCGCs increases as p increases. This distributions are calculated starting from $\tilde{Q} = 10^6$ pairs of random realizations of the initial damage.

existing among the state of different nodes (see Figure 7). We note that for $p < p_c$, in correspondence of the bimodal regime for the distribution $\pi(R)$, also $\rho(q)$ is bimodal, with a second peak at high values of the overlap that reflect configurations where the network is completely dismantled. Deviations of the distribution $\rho(q)$ from a single peak clearly indicate the complex many-body interactions present in the system, and reveals the important effect of the fluctuations of the MCGC in a way that is reminiscent of phenomena observed in disordered systems [42].

VI.   FINITE-SIZE EFFECTS

The discrepancy between $\hat{R}$ and $\bar{R}$ is an effect of the finite size of the networks analyzed. For an infinite network, the percolation transition is known to be self-averaging, i.e., the difference between $\bar{R}$ and $\hat{R}$ is vanishing. To explore for which network sizes we should expect significant differences between $\hat{R}$ and $\hat{R}$, we performed a

large deviation study of percolation on synthetic multiplex networks. We considered duplex networks of sizes $N = 10^2, 10^3, 10^4$ in which each layer is a random network with Poisson degree distribution and average degree $z = 5$. We observe that, as N increases, the percolation transition becomes self-averaging, and that $\bar{R}$ approximates increasingly better $\hat{R}$ (see Figure 8) , and the distribution $\pi(R)$ becomes increasingly narrow (see Supplementary Information [41]). Therefore, the distribution $\pi(R)$ is bimodal for a range of values of p that is converging to zero as N increases. Characterizing the large deviation of percolation is particularly important for investigating the robustness of small/medium size multiplex networks as the ones considered in this study. Interestingly, the average response to damage $\bar{R}$ and the typical response to damage $\hat{R}$ differ significantly up to network sizes of several thousand of nodes. Duplex networks of these size are very common, and include not only brain networks and air transportation networks such as those studied here, but also interdependent power-grids, ecological multiplex networks and brain functional networks. We believe therefore that our results might be
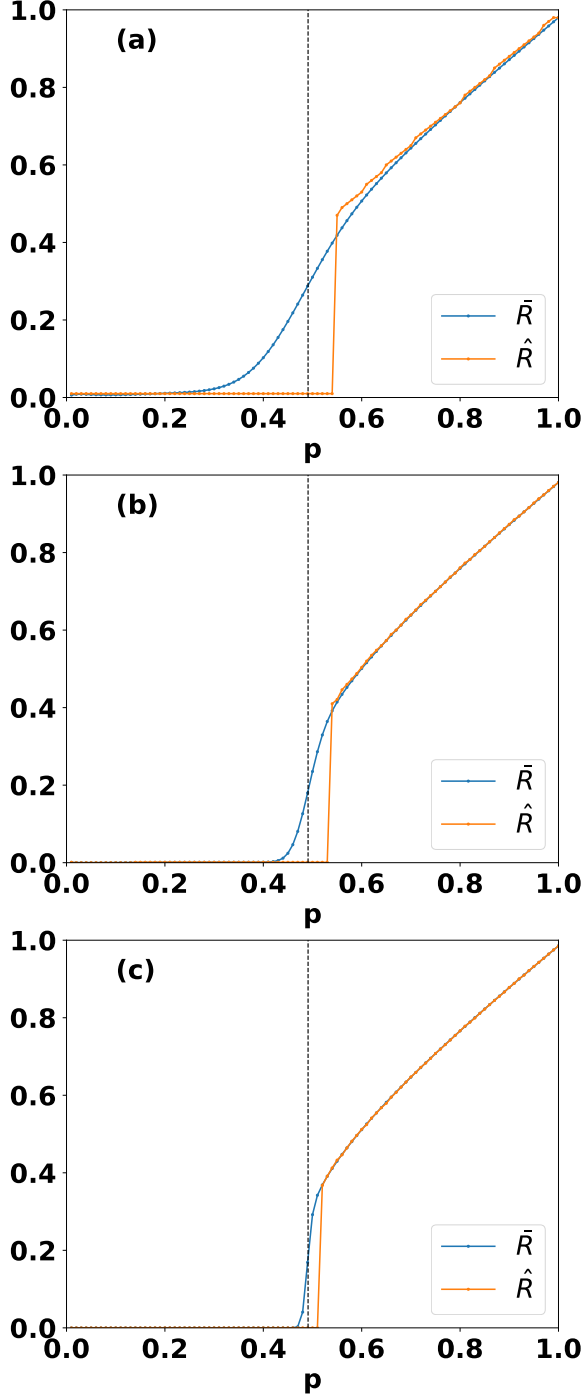
FIG. 8: **Finite-size effects of percolation on a Poisson duplex network.** The distribution $\pi(R)$ of the size of the MCGC, the average $\bar{R}$ and the typical $\hat{R}$ size of the MCGC are plotted versus p for Poisson duplex networks with average degree $z = 5$ and network sizes $N = 10^2$ (panel a), $N = 10^3$ (panel b), $N = 10^4$ (panel c). This numerical results are obtained by running $Q = 10^6$ (panel a,b) and $Q = 10^5$ (panel c) realizations of the initial damage configurations. The dashed solid line indicates the theoretically predicted percolation threshold in the limit $N \to \infty$.

relevant for scientists investigating the robustness of very different types of real multiplex datasets.

VII. SAFEGUARDING THE MCGC

Above, we defined the critical point p_c as the value of p where the two peaks of the bimodal distribution $\pi(R)$ have the same height. For $p = p_c$, we have that the left peak is located at $R = 1/N$, while the right peak is located at $R = R_c$, with $R_c \gg 1/N$. The condition $\pi(R = 1/N) = \pi(R = R_c)$ tells us that the likelihood that the system fails is comparable with the probability that the system is still in the functioning state. Is it possible to predict initial configurations of damage that lead to one or the other final states of the system? Is it possible to safeguard some nodes so that the sufficient condition that the network will be in the functional state is met? Please note that the latter question is different from the one defined in optimal percolation, where the goal is to dismantle a system rather than preserving its cohesiveness [33, 37].

Here, we propose an algorithm that ranks nodes according to their influence in determining the size of the MCGC. The algorithm uses the bimodality of $\pi(R)$, and is designed to be effective for $p = p_c$. We name the score resulting from the algorithm as *safeguard centrality*. The algorithm starts by defining two ranges of possible sizes for the MCGCs, either a well-defined MCGC $R > R^\star = \frac{1}{\sqrt{N}}$ or a dismantled network with $R < R^\star$. A score Δs_i is assigned to every node i. Δs_i is defined as the difference between the joint probability that node i is not initially damaged and $R > R^\star$, and the joint probability that node is not initially damaged and $R < R^\star$, i.e.,

$$\Delta s_i = \frac{1}{P} \sum_{\mu=1}^{P} s_i^\mu [\theta(R^\mu - R^\star) - \theta(R^\star - R^\mu)], \quad (13)$$

where $\theta(x)$ is the Heaviside function, i.e., $\theta(x) = 0$ if $x < 0$, and $\theta(x) = 1$, otherwise. Nodes with top Δs values are nodes whose safeguard may result in large MCGC sizes, i.e., the nodes that are responsible for keeping the multiplex connected.

In Figure 9a, we display Δs for each node of the American Airlines-United Airlines duplex network. The score seems informative. If the top-ranked nodes are damaged deterministically (see Figure 9b), the distribution $\pi(R)$ of the size of the giant component becomes more peaked around the value $R = 1/N$. If the top-ranked nodes are safeguarded (see Figure 9c), the robustness of the entire system in greatly improved. In fact, safeguarding the top-ranked node (Chicago O'Hare Airport, ORD) only is already sufficient to observe a unimodal distribution $\pi(R)$ with peak at $R = 0.3 > R_c$.

In order to investigate whether the top-ranked nodes according to Δs have some relation with the nodes identified in solutions to the optimal percolation problem,

Airport	Δs	SA	HDp	HDs	HDAp	HDAs	CI1p	CI1s
ORD	0.3984	0	1	1	1	1	1	1
DFW	0.3644	1	3	2	3	2	8	6
LAX	0.3523	0	2	5	2	5	2	5
MIA	0.3462	1	7	7	8	7	9	9
SFO	0.3273	0	4	6	5	6	3	4
Kendall-τ	1	0.8	0.6	0.8	0.6	0.8	0.4	0.2

TABLE II: The top 5 airports in the American Airlines-United Airlines duplex network according to the centrality measure Δs are listed together with their corresponding classification $\{s_i^\star\}$ according to the SA algorithm ($s_i^\star = 0$, if node i belongs to the set of structural nodes, $s_i^\star = 1$, otherwise) and their rank according to the HDp, HDs, HDAp, HDAs, CI1p and CI1s algorithms. The last row indicates the Kendall-τ correlations among the ranking of these 5 airports according to Δs and each of the other state of the art algorithms. Note that when comparing to the SA results we have used the Kendall τ-c [43] correlation coefficient while we have used the Kendall τ-a [44] correlation coefficient in all the other cases.

we performed systematic comparisons between the top 5 nodes according to Δs and various methods used in optimal percolation [37]. We find that the top-ranked nodes correspond with good accuracy to the nodes in the optimal structural set detected by Simulated Annealing optimization [37]. High correlation (measured using Kendall τ) is also found with sets determined using other state-of-the-art techniques (see Table II). These include the High Degree (HD) and the High Degree Adaptive (HDA) algorithms based on the product (HDp,HDAp) or the sum (HDs,HDAs) of the node degree in the two layers, and the duplex network version of the Collective Influence (CI) algorithm based on the product (CIℓp) or on the sum (CIℓs) of the CI scores of single layers (see Supplementary Information for details). In the Supplementary Information, we present the same type of analysis for for the United Airlines-Delta Airlines duplex network yielding similar conclusions.

VIII. CONCLUSIONS

We explored the large deviation properties of percolation of real finite multiplex networks. This approach consists in looking at the distribution $\pi(R)$ of the size R of the MCGC as a function of the probability p that a node is not initially damaged. The motivation of the study finds its roots from the obvious inability of the mean value $\bar{R}$ to capture large fluctuations that arise in finite-size systems. Although the use of $\pi(R)$ is required to fully characterize the properties of the percolation transition in real multiplex networks, in the paper, we demonstrated that most of the system robustness can be understood by combining the information provided already by $\bar{R}$ with the complementary metric $\hat{R}$, i.e., the mode of the distribution $\pi(R)$. $\hat{R}$ reveals the intrinsic fragility of a real multiplex displaying a discontinuity as a function of the probability p that a node is not initially damaged. This discontinuity characterizes the position of an effective critical point $p = p_c$ where the distribution $\pi(R)$ of the sizes R of the MCGC is bimodal and displays two local maxima of the same height at $R = 1/N$ (indicating that the network is totally dismantled) and at $R = R_c \gg 1/N$ (indicating that the network has a significantly large MCGC). Therefore, for $p = p_c$, the possible outcome of an initial damage is very uncertain.

The large deviation approach to percolation allows us to characterize the correlations among the state of different nodes in the network and the fluctuations in the state of single nodes measured by the so called specific heat of percolation. Note that here we indicate by state of a node its inclusion or exclusion from the MCGC resulting from a given realization of the initial damage. We show that nearest neighbor nodes display an average correlation that has a maximum for a value of p close to the percolation threshold p_c.

Finally, we focused our attention on the destiny of the MCGC at $p = p_c$ proposing an algorithm able to detect some special nodes. The safeguard of these nodes can ensure with high probability that the most likely outcome is $\hat{R} > R_c$ and that the total dismantling of the network has a suppressed probability. The proposed algorithm was tested on real datasets showing the efficiency of the proposed safeguarding procedure. We further showed the set of top-scoring nodes is almost identical to those found as solutions to the optimal percolation problem.
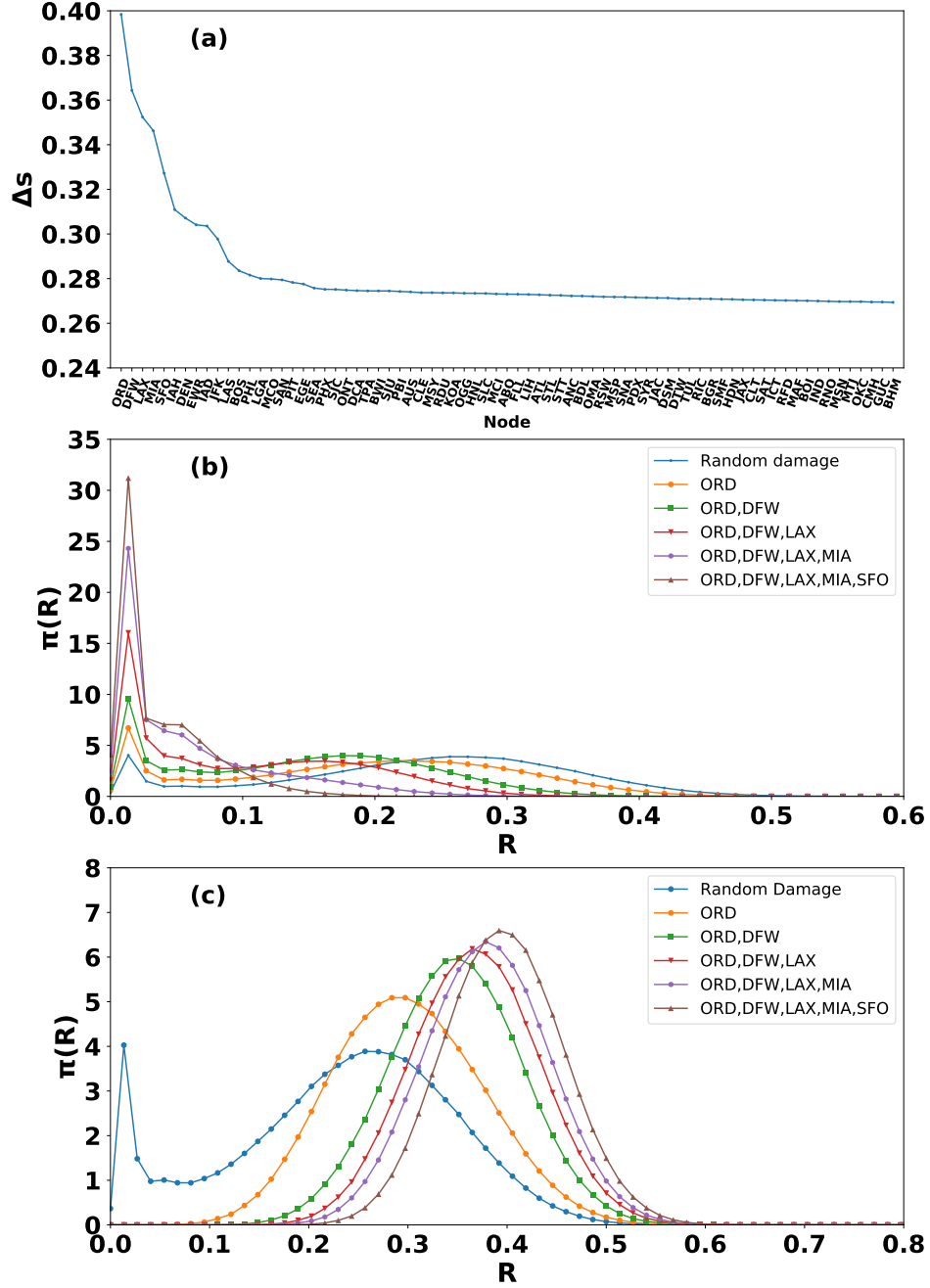
FIG. 9: **Effect of the damage and the safeguard of the top-ranked nodes on the robustness of the American Airlines-United Airlines duplex network.** The centrality measure Δs for each of the $N = 73$ airports of the American Airlines-United Airlines dataset is shown in panel (a). The centrality measures is evaluated by considering $P = 10^6$ realization of the initial damage at $p = p_c = 0.40$ taking $R^\star = 1/\sqrt{N} < R_c = 0.27$. The distribution $\pi(R)$ of the size R of the MCGC at $p = p_c$ is compared to the distribution $\pi(R)$ obtained when the top-ranked nodes according to Δs are damaged for sure (panel (b)) or safeguarded (panel (c)) while the other nodes are damaged with probability $p = p_c = 0.40$. The distribution in panels (b) and (c) are obtained considering 10^6 realizations of the initial damage.

[1] G. Bianconi, *Multilayer Networks: Structure and Function* (Oxford University Press, Oxford, 2018).

[2] S. Boccaletti et al., The structure and dynamics of multilayer networks. *Physics Reports* **544**, 1 (2014).

[3] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno and M.A. Porter, Multilayer networks. *Jour. Comp. Net.* **2**, 203 (2014).

[4] K. M. Lee, B. Min and K.I. Goh, Towards real-world complexity: an introduction to multiplex networks. *Eur. Phys. Jour. B* **88**, 48 (2015).

[5] G. Bianconi, Statistical mechanics of multiplex networks: Entropy and overlap. *Phys. Rev. E* **87**, 062806 (2013).

[6] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley and S. Havlin, Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025 (2010).

[7] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev adn J. F. F. Mendes, Avalanche collapse of interdependent networks. *Phys. Rev. Lett.* **109**, 248701 (2012).

[8] B. Min, S. D. Yi, K.-M. Lee and K.-I. Goh, Network robustness of multiplex networks with interlayer degree correlations. *Phys. Rev. E* **89**, 042811 (2014).

[9] S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, M. Paczuski, Percolation theory on interdependent networks based on epidemic spreading. *EPL (Europhysics Letters)* **97**, 16006 (2012).

[10] R. Parshani, C. Rozenblat, D. Ietri, C. Ducruet and S. Havlin, Inter-similarity between coupled networks. *EPL (Europhysics Letters)* **92**, 68002 (2010).

[11] R. Parshani, S. V. Buldyrev and S. Havlin, Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. *Phys. Rev. Lett.* **105**, 048701 (2010).

[12] G. Bianconi, S. N. Dorogovtsev and J. F. F. Mendes, Mutually connected component of networks of networks with replica nodes. *Phys. Rev. E* **91**, 012804 (2015).

[13] G. Bianconi and S. N. Dorogovtsev, Multiple percolation transitions in a configuration model of a network of networks. *Phys. Rev. E* **89**, 062814 (2014).

[14] F. Radicchi and G. Bianconi, Redundant interdependencies boost the robustness of multiplex networks. *Phys. Rev. X* **7**, 011013 (2017).

[15] G. Bianconi and F. Radicchi, Percolation in real multiplex networks. *Phys. Rev. E* **94**, 060301 (2016).

[16] D. Cellai, S. N. Dorogovtsev and G. Bianconi, Message passing theory for percolation models on multiplex networks with link overlap. *Phys. Rev. E* **94**, 032301 (2016).

[17] G. J. Baxter, G. Bianconi, R. A. da Costa, S. N. Dorogovtsev and J. F. F. Mendes, Correlated link overlaps in Multiplex Networks. *Phys. Rev. E* **94**, 012303 (2016).

[18] D. Cellai, E. López, J. Zhou, J. P. Gleeson and G. Bianconi, Percolation in multiplex networks with overlap. *Phys. Rev. E* **88**, 052811 (2013).

[19] F. Radicchi, Percolation in real interdependent networks. *Nature Phys.* **11**, 597 (2015).

[20] B. Min, S. Lee, K.-M. Lee and K.I. Goh, Link overlap, viability, and mutual percolation in multiplex networks. *Chaos, Solitons Fractals* **72** 49 (2015).

[21] S. D. S. Reis, Y. Hu, A. Babino, J. S. Andrade Jr., S. Canals, M. Sigman and H. A. Makse, Avoiding catastrophic failure in correlated networks of networks. *Nature Phys.* **10**, 762 (2014).

[22] R. D'Souza and J. Nagler, Anomalous critical and supercritical phenomena in explosive percolation. *Nature Physics* **11**, 531 (2015).

[23] J. Gómez-Gardeñes, S. Gómez, A. Arenas and Y. Moreno, Explosive synchronization transitions in scale-free networks. Phys. Rev. Lett. **106**, 128701 (2011).

[24] V. Nicosia, P. S. Skardal, A. Arenas and V. Latora, Collective phenomena emerging from the interactions between dynamical processes in multiplex networks. *Phys. Rev. Lett.* **118**, 138302 (2017).

[25] F. Radicchi and A. Arenas, Abrupt transition in the structural formation of interconnected networks. *Nature Phys.* **9**, 717 (2013).

[26] S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, K-core organization of complex networks. *Phys. Rev. Lett.* **96**, 040601 (2006).

[27] G. Parisi and T. Rizzo, k-core percolation in four dimensions. *Phys. Rev. E* **78**, p.022101 (2008).

[28] S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, Critical phenomena in complex networks. Rev. Mod. Phys. **80**, 1275 (2008).

[29] B. Karrer, M. E. J. Newman and L. Zdeborová, Percolation on sparse networks. *Phys. Rev. Lett.* **113**, 208702 (2014).

[30] W. K. Chai, V. Kyritsis, K. V. Katsaros and G. Pavlou, Resilience of interdependent communication and power distribution networks against cascading failures. In IFIP Networking Conference (IFIP Networking) and Workshops, 2016 (pp. 37-45). IEEE (2016).

[31] G. Bianconi, Fluctuations in percolation of sparse complex networks. *Phys. Rev. E* **96**, 012302 (2017).

[32] G. Bianconi, Rare events and discontinuous percolation transitions. *Phys. Rev. E* **97**, 022314 (2018)

[33] F. Morone and H.A. Makse, Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65 (2015).

[34] A. Braunstein, L. DallAsta, G. Semerjian and L. Zdeborová, L. Network dismantling. *Proc. Nat. Aca. Sci.* **113**, 12368 (2016).

[35] D. Lee, S. Choi, M. Stippinger, J. Kertész, and B. Kahng, Hybrid phase transition into an absorbing state: Percolation and avalanches.*Phys. Rev. E* **93**, 042109 (2016).

[36] M. Kitsak, A. A. Ganin, D. A. Eisenberg, P. L. Krapivsky, D. Krioukov, D. L. Alderson and I. Linkov, Stability of a giant connected component in a complex network. *Phys. Rev. E* **97**, 012309 (2018).

[37] S. Osat, A. Faqeeh and F. Radicchi, Optimal percolation on multiplex networks. *Nat. Comm.* **8**, 1540 (2017).

[38] G. J. Baxter, G. Timár and J. F. F. Mendes, Targeted Damage to Interdependent Networks. arXiv preprint arXiv:1802.03992 (2018).

[39] B. L. Chen, BD. H. Hall and D. B. Chklovskii, Wiring optimization can relate neuronal structure and function" *Proc. Nat. Aca. Sci.* **103**, 4723 (2006)

[40] M. De Domenico, M.A. Porter and A. Arenas, MuxViz: A Tool for Multilayer Analysis and Visualization of Networks. *Jour. Comp. Net.* **3**, 159 (2015).

[41] Supplementary Information available at

[42] M. Mézard, G. Parisi, and M. Virasoro, *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications* (World Scientific Publishing Company

1987).

[43] K. J. Berry, J. E. Johnston, S. Zahran, P. W. Mielke, Stuart's tau measure of effect size for ordinal variables: Some methodological considerations. *Behavior Research Methods* **41**, 1144 (2009).

[44] M. A. Kendall, New Measure of Rank Correlation. *Biometrika* **30**, 81 (1938).