Tuning the overlap and the cross-layer correlations in two-layer networks: Application to a susceptible-infectious-recovered model with awareness dissemination

David Juher[*]
Department of IMAE, Universitat de Girona, Girona 17003, Catalonia, Spain

Joan Saldaña[†]
Department of IMAE, Universitat de Girona, Girona 17703, Catalonia, Spain
and Department of Electrical and Computer Engineering, Kansas State University, Manhattan, Kansas 66506, USA

(Received 29 November 2017; published 9 March 2018)

We study the properties of the potential overlap between two networks A, B sharing the same set of N nodes (a two-layer network) whose respective degree distributions $p_A(k), p_B(k)$ are given. Defining the overlap coefficient α as the Jaccard index, we prove that α is very close to 0 when A and B are random and independently generated. We derive an upper bound α_M for the maximum overlap coefficient permitted in terms of $p_A(k)$, $p_B(k)$, and N. Then we present an algorithm based on cross rewiring of links to obtain a two-layer network with any prescribed α inside the range $(0,\alpha_M)$. A refined version of the algorithm allows us to minimize the cross-layer correlations that unavoidably appear for values of α beyond a critical overlap $\alpha_c < \alpha_M$. Finally, we present a very simple example of a susceptible-infectious-recovered epidemic model with information dissemination and use the algorithms to determine the impact of the overlap on the final outbreak size predicted by the model.

DOI: 10.1103/PhysRevE.97.032303

I. INTRODUCTION

Some contagious processes interact with each other during their propagation, which can occur either through the same route of transmission or through routes that share the same set of nodes but use different types of connections. In the second case, the description of the spread uses the concept of multilayer or multiplex network, namely, a set of nodes (individuals, computers, etc.) connected by qualitatively different types of links corresponding to possible relationships among them (acquaintanceship, friendship, physical contact, social networks, etc.), each layer defined by a type of connection. Competitive viruses spreading simultaneously through different routes of transmission over the same host population, or the spread of a pathogen and awareness during an epidemic episode are examples of processes that are better described by means of multilayer networks [1].

In the last years it has been a development of the mathematical formulation of multiplex networks and, also, of more general interconnected networks for which the set of nodes does not need to be the same at each layer [2–4]. Moreover, recent results show the importance of the interrelation between different layers in determining the fate of competitive epidemic processes [1,5]. In other cases, however, the importance of such an interrelation is not so evident from the analytical results of the epidemic threshold [6,7], or even seems to be not relevant at all [8].

Only a few papers dealing with competing epidemics over multilayer networks focus on the impact of layer overlap on the epidemic dynamics [5,9,10]. In [5], the authors consider a sequential propagation of two epidemics using distinct routes of transmission over a network consisting of two partly overlapped layers. Using bond percolation, it is determined the success of a second epidemic through that part of its route of transmission whose nodes have not been infected by the first epidemic. In [10], the authors develop an analytical approach to deal with simultaneous spread of two interacting viral agents on two-layered networks. In that work, moreover, the respective effects of overlap and correlation of the degrees of nodes in each layer on the epidemic dynamics are considered.

Here the *overlap* α between two (labeled) networks A and B of N nodes is defined as the fraction of links of the union network that are common links of A and B or, equivalently, the probability that a randomly chosen link of the network $A \cup B$ is simultaneously a link of both A and B. In fact α is the *Jaccard index*, a statistic used for comparing the similarity of two sample sets, as defined in [11]. Just to illustrate that this simple statistical parameter can play a critical role in the qualitative response of a two-layer network model, in Sec. VIII we present a mean-field model for the spread of an infectious agent on one layer (contact layer). The model implicitly assumes an information dissemination on a second layer (notification layer) about the infection status of the nodes which causes an increase in awareness and the adoption of preventive behaviors. As an interesting feature, the overlap coefficient α between the networks embedding the respective routes of transmission is a parameter of the model. This allows us to derive a simple relationship between α and the epidemic threshold. Provided that one wants to perform simulations to validate this (or any) model, a systematic procedure to generate couples of networks of given size and degree distributions with a prescribed value of α would be a useful tool. We stress

[*]david.juher@udg.edu
[†]joan.saldana@udg.edu

that this is the main focus of the paper, and that the model in Sec. VIII is just a simple example to illustrate the convenience of having such tools.

Our approach is based on the study of the potential overlap between two networks whose (finite, empirical) degree distributions are previously fixed. More precisely, in Secs. III and IV we estimate the minimum and maximum values (call them α_m and α_M) for the overlap coefficient between two networks of size N and degree distributions $p_A(k)$ and $p_B(k)$. In particular, we show that $\alpha_m \approx 0$. The study of the maximum α_M is based on the computation of the *potential* overlap $\mathcal{O}^{\mathrm{pot}}(D_A,D_B)$ between two fixed degree sequences D_A,D_B following $p_A(k)$, $p_B(k)$. In Sec. V we present the *CR algorithm*, that takes as input any two degree sequences D_A,D_B and a desired overlap $\alpha \in (0,\mathcal{O}^{\mathrm{pot}}(D_A,D_B))$ and generates a couple of networks with degree sequences D_A,D_B and overlap coefficient close to α. When D_A,D_B are *randomly* sampled from $p_A(k),p_B(k)$, the potential overlap $\mathcal{O}^{\mathrm{pot}}(D_A,D_B)$ is called *critical overlap*. In Sec. VI we show that the CR algorithm starting with two random sequences succeeds in constructing pairs of networks having any overlap below the critical one and exhibiting some desirable statistical properties, specifically lack of in- and cross-layer degree-degree correlations. Of course the critical overlap belongs to the interval (α_m,α_M), and is higher than expected from intuition. In Sec. VII we show that for values between the critical overlap and α_M there is an unavoidable direct relationship between overlap and cross-layer degree-degree correlation, and propose a refined version of the cross-rewiring (CR) algorithm that tries to reach values beyond the critical overlap while maintaining the cross-layer degree-degree correlation as small as possible.

With this collection of algorithms, we are given a tool to test the analytical predictions relating overlap and epidemic thresholds. In the few previous works dealing with interacting epidemics on overlay networks [5,9,10], the two-layer network over which multiple pathogens spread was characterized by the probability $\rho(k_A,k_B,k_c)$ that a randomly selected node belongs to k_A links unique to layer A, k_B links unique to layer B, and k_c common links [note that, using this language, what we are assuming here as a natural requirement is that the *marginal* distributions $p_A(k_A)$ and $p_B(k_B)$, that can be recovered as $p_A(k_A)=\sum_{k_B,k_c}\rho(k_A-k_c,k_B-k_c,k_c)$ and $p_B(k_B)=\sum_{k_A,k_c}\rho(k_A-k_c,k_B-k_c,k_c)$, are given]. Those papers put the main focus on the influence of the overlap and the degree correlations on the epidemic dynamics predicted by the model, rather than on the algorithms used to construct the two-layer network. So, the simulations to test the validity of the predictions were performed over particularly simple cases (rich enough, nevertheless, to extract valid conclusions). As an example, to test the model response to an arbitrary overlap α, the authors perform the simulations in the simplest setting $p_A(k)=p_B(k)\equiv p(k)$, that obviously admits any overlap coefficient from 0 to 1, take

$$\rho(k_A,k_B,k_c)=p(k_A+k_c)\delta_{k_A,k_B}\binom{k_A+k_c}{k_c}\alpha^{k_c}(1-\alpha)^{k_A}$$

(attach independently $k_A+k_c=k_B+k_c$ links to a node, with a probability α for each link to belong to both networks), and execute a "configuration model"-like algorithm to connect pairs of stubs sampled from ρ with the obvious restrictions. As another example, the respective effects of overlap and degree correlations are isolated by considering again simple (and extremal) cases: random overlap and no degree correlation; random overlap and full degree correlation; and full overlap.

In a more general setting ($p_A \neq p_B$), it is not straightforward to extend the configuration model algorithm to get prescribed intermediate overlaps and/or degree correlations. In contrast, the algorithm we present here can be used to generate any permitted value of both parameters in the range forced by the marginal degree distributions.

II. TERMINOLOGY AND STANDING NOTATION

Throughout this paper, the nodes of any network will be labeled with the natural numbers $\{1,2,\ldots,N\}$. The cardinality of a finite set X will be denoted by $|X|$. Let $V=\{1,2,\ldots,N\}$ for some $N\in\mathbb{N}$. Let E and E' be two subsets of $\{\{i,j\}: i\neq j \text{ and } i,j\in V\}$. Let G and G' be the undirected networks having V as the set of nodes and E and E' as the respective sets of links. The *union network* $G\cup G'$ is the undirected network whose sets of nodes and links are V and $E\cup E'$ respectively. By definition, we will say that G and G' are *different* from each other if and only if $E\neq E'$. In particular, if we have a network H and we simply permute the labels of the nodes of H, then we obtain a network that is in general different from (but isomorphic to) H. Observe that the union operation is not a topological invariant: the union of two networks does not depend only on their shapes but also on the way their nodes are labeled. The *overlap* between G and G' is defined as the fraction

$$\mathcal{O}(G,G'):=\frac{|E\cap E'|}{|E\cup E'|}=\frac{|E\cap E'|}{|E|+|E'|-|E\cap E'|},$$

which can be thought of as the probability that a randomly chosen link of $G\cup G'$ is simultaneously a link of both G and G'.

A *degree set* of cardinality N is a multiset (i.e., multiple instances of each element are allowed) of N integers that is realizable as the set of degrees of a network. That is, there exist a labeling $\{k_1,k_2,\ldots,k_N\}$ of the elements of the set and a network G of N nodes such that k_i is the degree of the node i. Equivalently, $\sum k_i$ is even and the integers k_i satisfy the well-known Havel-Hakimi condition [12,13] (a technical recursive condition that is irrelevant to our purposes). As usual, the ordered list $D=(k_1,k_2,\ldots,k_N)$ will be called the *degree sequence* of G. Note that rearranging the elements of D by means of a permutation σ corresponds to relabeling the nodes of G to get another network G' isomorphic to G with degree sequence $\sigma(D)$ (and the same degree set as G).

A probability distribution $p(k)$ with bounded support will be called *empirical (of N nodes)* if it is realizable as the degree distribution of a network of N nodes. That is, there exists a network G of N nodes such that, if $\{k_1,k_2,\ldots,k_N\}$ is the degree set of G, then $N_k:=|\{i:k_i=k\}|=p(k)N$. Observe that giving an empirical distribution $p(k)$ of N nodes is completely equivalent to specify a degree set K of cardinality N. For an ordered sequence D we will write $D\sim p(k)$ to indicate that D is a particular arrangement of the elements of K. For a pair of ordered sequences (D,D') and two empirical distributions $p(k),p'(k)$, we will write $(D,D')\sim p(k)\times p'(k)$ to indicate that $D\sim p(k)$ and $D'\sim p'(k)$.

We use the term *empirical* for a degree distribution to distinguish it from a (theoretical, not necessarily with bounded support) probability distribution $p(k)$. In this case, for any $N \in \mathbb{N}$, one can use several standard algorithms (see Sec. III) to construct a network G_N of N nodes whose empirical degree distribution $p_N(k)$ is close to $p(k)$, in the sense that, for big enough values of N, $p_N(k)$ converges in probability to $p(k)$ ([14], Theorem 2.1).

Assume that we are given two empirical degree distributions $p(k), p'(k)$ of N nodes, with corresponding degree sets K and K'. Let n and n' be the total number of pairwise different networks having respectively K and K' as degree sets, each one numbered with an integer in the range $[1,n]$ (respectively, $[1,n']$). Then we can clearly consider a function of two variables $\mathcal{O}(x,y)$ on the grid of all pairs (x,y) of integers in $[1,n] \times [1,n']$, that gives the value of the overlap of the networks numbered as x and y. Observe that the function $\mathcal{O}(x,y)$ has a global minimum and maximum. These extremal values will be denoted by $\underline{\mathcal{O}}(N,p,p')$ and $\overline{\mathcal{O}}(N,p,p')$, or simply by $\underline{\mathcal{O}}$ and $\overline{\mathcal{O}}$ when no confusion seems possible.

III. EXPECTED OVERLAP BETWEEN TWO RANDOM INDEPENDENT LAYERS

Assume that we are given two empirical degree distributions $p(k), p'(k)$ of N nodes. In this section we prove that the expected overlap between two random networks of N nodes and degree distributions $p(k)$ and $p'(k)$ (generated, for instance, via the standard *configuration model algorithm* [15–17]) is very close to zero when N is big enough, thus showing that $\underline{\mathcal{O}} \approx 0$. Giving estimations for $\overline{\mathcal{O}}$ will be the matter of Sec. IV.

Let us recall the configuration model algorithm to generate a random network with a given degree sequence $(k_1,k_2,\ldots,k_N)$. Take a vector X of length $2L := \sum k_i$ containing k_1 times the integer 1 in the first k_1 entries, k_2 times the integer 2 in the following k_2 entries, etc. Each entry v of X represents a single stub (or semilink) attached at the node labeled as v. Then, take a random permutation of the entries of X to get a new array Y. Finally, read the contents of Y in order, interpreting each pair of consecutive entries v,w as a link between the nodes v and w. For an example, take $N = 6$ and consider the degree distribution $p(k)$ defined by $p(1) = p(3) = 1/6$, $p(2) = 4/6$, and $p(k) = 0$ for $k \neq 1,2,3$. The corresponding degree set is $\{1,2,2,2,2,3\}$. Take, for instance, $(1,2,2,2,2,3)$ as degree sequence. Then, $X = (1,2,2,3,3,4,4,5,5,6,6,6)$. Now we permute X at random, obtaining $Y = (3,4,5,1,6,3,6,2,4,5,2,6)$. The links of the obtained network are $\{3,4\}$, $\{5,1\}$, $\{6,3\}$, $\{6,2\}$, $\{4,5\}$, $\{2,6\}$. Observe that the link $\{6,2\}$ appears twice. In general, the configuration model algorithm gives *multigraphs* rather than graphs. It is well known, however, that the fraction of self-loops and multilinks over the total number of links goes to 0 as $N \to \infty$ when the variance of the degree distribution is bounded [18]. See [14] for alternative implementations of the configuration model to get simple graphs.

It seems natural to expect that the overlap between two networks of respective degree distributions $p(k), p'(k)$ and size N generated via the configuration model algorithm is very small. When the respective mean degrees are small with respect to the total size N this turns out to be true. To prove this fact, we need to estimate the probability that two given nodes are connected in a random network generated via the configuration model algorithm. So, let G be a network of N nodes, L links, and degree distribution $p(k)$. Assume that G has been obtained by means of the configuration model algorithm starting with a degree sequence $(k_1,k_2,\ldots,k_N)$. Take at random any pair $\{i,j\}$ of nodes with $k_i \leqslant k_j$. Next we estimate the probability p_{ij} that the network G contains the link $\{i,j\}$. This probability is given by the quotient a/b, where b is the total number of rearrangements Y of the vector X (here we are using the notation introduced in the definition of the configuration model) and a is the number of such rearrangements having at least two consecutive entries i,j (or j,i) in places Y_n, Y_{n+1} for $n = 1,3,5,\ldots,2L-1$. We have that

$$b = \frac{(2L)!}{k_1!k_2!\cdots k_N!}. \tag{1}$$

Let us compute a. For $l = 1,2,\ldots,L$, let Y^l be the set of rearrangements Y containing the entries i,j (or j,i) in places Y_{2l-1}, Y_{2l}. Then, $a = |Y^1 \cup Y^2 \cup \ldots \cup Y^L|$. By the inclusion-exclusion principle, $a = a_1 - a_2 + \cdots + (-1)^{k_i-1} a_{k_i}$, where a_l is the sum of the cardinalities of all intersections of l sets in $Y^1, Y^2, \ldots, Y^L$. A simple combinatorial argument yields that, for $l \leqslant k_i$,

$$a_l = \frac{\binom{L}{l} 2^l (2L-2l)!}{k_1!\ldots k_{i-1}!(k_i-l)!k_{i+1}!\ldots k_{j-1}!(k_j-l)!k_j!\ldots k_N!},$$

while $a_l = 0$ for $k_i < l \leqslant L$. Using the previous expression and the inclusion-exclusion principle we get that

$$a = \frac{\sum_{l=1}^{k_i} (-1)^{l-1} \binom{L}{l} 2^l \frac{(2L-2l)!}{(k_i-l)!(k_j-l)!}}{k_1!k_2!\ldots k_{i-1}!k_{i+1}!\ldots k_{j-1}!(k_{j+1})!\ldots k_N!}.$$

Taking it all into account, we get that the probability that G contains the link $\{i,j\}$ is

$$p_{ij} = \frac{L!k_i!k_j!}{(2L)!} \sum_{l=1}^{k_i} \frac{(-1)^{l-1} 2^l (2L-2l)!}{l!(L-l)!(k_i-l)!(k_j-l)!}. \tag{2}$$

This exact expression is too complex to be used to estimate the expected overlap between two random networks. Instead, if in the previous proof we replace a simply by a_1, then it easily follows that

$$p_{ij} \approx \frac{k_i k_j}{2L-1}, \tag{3}$$

that is in fact a standard approximation used in the literature for the probability p_{ij} [18,19]. The approximation (3) is good enough only when k_i and k_j are small with respect to L, in particular when we consider networks with bounded mean degree and large size N, which is the case for most modeling applications. However, in general (3) can significantly differ from the exact formula (2).

Now let $p(k), p'(k)$ be two empirical degree distributions with respective means $\langle k \rangle$ and $\langle k' \rangle$. Let G, G' be two networks of N nodes and degree distributions $p(k)$ and $p'(k)$ generated via the configuration model algorithm starting with degree sequences $(k_1,k_2,\ldots,k_N)$ and $(k'_1,k'_2,\ldots,k'_N)$. Assume that N is big enough with respect to $\langle k \rangle$ and $\langle k' \rangle$ in such a way that the approximation (3) holds. Let L, L' be the number of links of G and G' respectively. Using (3) we can compute the probability

p that two different nodes chosen at random are neighbors in G:

$$p \approx \frac{1}{2L-1}\sum_{k_i,k_j} k_i\, p(k_i) k_j\, p(k_j) = \frac{\langle k\rangle^2}{2L-1} \approx \frac{\langle k\rangle}{N}, \quad (4)$$

where in the last expression $\langle k\rangle$ denotes the expected degree of a node and we have used that $\langle k\rangle N = 2L$. Now the expected overlap between G and G' can be computed as the probability that two different nodes are connected in both G and G' over the probability that they are connected in $G \cup G'$ which, by virtue of (4), is

$$\frac{\langle k\rangle\langle k'\rangle/N^2}{1-\left(1-\frac{\langle k\rangle}{N}\right)\left(1-\frac{\langle k'\rangle}{N}\right)}.$$

In consequence,

$$\mathcal{O}(G,G') \approx \frac{\langle k\rangle\langle k'\rangle}{N(\langle k\rangle+\langle k'\rangle)-\langle k\rangle\langle k'\rangle}, \quad (5)$$

telling us that, given N and any two degree distributions $p(k), p'(k)$, the minimum overlap $\underline{\mathcal{O}}(N,p,p')$ is very close to 0, at least when N is big with respect to the expected values $\langle k\rangle$ and $\langle k'\rangle$. Of course, for small networks this is not true in general.

IV. AN UPPER BOUND FOR THE MAXIMUM OVERLAP

We start this section by giving a computable upper bound for $\overline{\mathcal{O}}(N,p,p')$ in terms of the size N and the empirical distributions $p(k), p'(k)$. To do it, first we introduce the notion of *potential overlap* between two *fixed* degree sequences.

Let G,G' be two networks of N nodes and empirical degree distributions $p(k), p'(k)$, with means $\langle k\rangle$ and $\langle k'\rangle$ and corresponding degree sequences $D = (k_1,k_2,\ldots,k_N)$, and $D' = (k'_1,k'_2,\ldots,k'_N)$, with $\sum k_i = \langle k\rangle N =: 2L$ and $\sum k'_i = \langle k'\rangle N =: 2L'$. If E and E' are the sets of links of G and G', then by definition

$$\mathcal{O}(G,G') = \frac{|E\cap E'|}{|E\cup E'|} = \frac{|E\cap E'|}{L+L'-|E\cap E'|} = \frac{x}{(\langle k\rangle+\langle k'\rangle)\frac{N}{2}-x} =: F(x), \quad (6)$$

where x stands for $|E\cap E'|$. Now observe that $F(x)$ is increasing in x. In consequence, an upper bound for the overlap is obtained when replacing x by the maximum possible number of links of the intersection network. It is clear that the intersection network cannot have more than $\min\{k_i,k'_i\}$ links attached at node i. In consequence, the total number of links of the intersection network is at most

$$\frac{1}{2}\sum_{i=1}^{N}\min\{k_i,k'_i\}.$$

So, we define the *potential overlap* $\mathcal{O}^{\rm pot}(D,D')$ associated to a pair (D,D') of degree sequences as

$$\frac{\frac{1}{2}\sum_i \min\{k_i,k'_i\}}{\frac{1}{2}\sum_i(k_i+k'_i)-\frac{1}{2}\sum_i\min\{k_i,k'_i\}}, \quad (7)$$

TABLE I. Critical overlap as defined in Sec. VI (first row) and the upper bound (9) for the maximum overlap permitted (second row) between pairs of empirical distributions. In all cases $N = 10\,000$. For the left column distributions, $\langle k\rangle = 20$ while, for the upper ones, $\langle k\rangle = 26$.

	Regular	Poisson	SF	Exponential
	0.7693	0.7508	0.6301	0.6654
Regular	0.7693	0.7508	0.6301	0.6654
	0.7552	0.7259	0.5969	0.6392
Poisson	0.7552	0.7709	0.7221	0.7739
	0.5451	0.5365	0.4903	0.5117
SF	0.5451	0.6000	0.7688	0.7023
	0.6330	0.6174	0.5415	0.5683
Exponential	0.6330	0.7077	0.7715	0.7706

that, since $k_i + k'_i = \max\{k_i,k'_i\} + \min\{k_i,k'_i\}$, can be rewritten as

$$\mathcal{O}^{\rm pot}(D,D') := \sum_{i=1}^{N}\min\{k_i,k'_i\} \Big/ \sum_{i=1}^{N}\max\{k_i,k'_i\}. \quad (8)$$

Now observe that

$$\overline{\mathcal{O}} \leqslant \max_{(D,D')\sim p\times p'}\{\mathcal{O}^{\rm pot}(D,D')\}$$

and recall that the set of possible degree sequences associated to $p(k)$ coincides essentially with the set of all permutations of the numbers $k_1,k_2,\ldots,k_N$. Thus, if $(D,D') \sim p(k)\times p'(k)$ and σ,ρ are two permutations of order N, then $(\sigma(D),\rho(D')) \sim p(k)\times p'(k)$. Moreover, $\mathcal{O}^{\rm pot}(\sigma(D),\sigma(D')) = \mathcal{O}^{\rm pot}(D,D')$. In consequence, without loss of generality we can assume that D is increasingly ordered (that is, $k_i \leqslant k_j$ if $i<j$). In this case, it is easy to check that if there is a pair of entries $k'_i \geqslant k'_j$ of D' with $i<j$, then if we swap both entries the obtained sequence D'' satisfies $\mathcal{O}^{\rm pot}(D,D') \leqslant \mathcal{O}^{\rm pot}(D,D'')$. So, the maximum in the previous inequality is attained precisely when both D and D' are increasingly ordered. So, we have proved that

$$\overline{\mathcal{O}} \leqslant \frac{\sum_{i=1}^{N}\min\{k_i,k'_i\}}{\sum_{i=1}^{N}\max\{k_i,k'_i\}}, \text{ whenever}$$
$$k_1 \leqslant k_2 \leqslant \cdots \leqslant k_N \text{ and } k'_1 \leqslant k'_2 \leqslant \cdots \leqslant k'_N. \quad (9)$$

Inequality (9) allows us to design an efficient algorithm to compute an upper bound for the maximum overlap. The algorithm takes as input the empirical distributions $p(k)$ and $p'(k)$, sorts increasingly the elements of the respective degree sets, and finally returns the right-hand side of the inequality in (9). Table I (second row in each box) shows the output of this algorithm for several pairs of empirical distributions, obtained by approximating the corresponding pairs of (theoretical) distributions. Here "SF" stands for a scale-free network with $p(k) = Ck^{-\gamma}$ with $\gamma = 3$, minimum degree m, cutoff $k_c = mN^{1/2}$, and the normalization constant $C = (\gamma-1)m^{\gamma-1}N/(N-1)$, for which $\langle k\rangle \approx 2m$ [20]. "Exponential" corresponds to $p(k) = (1/m)e^{1-k/m}$ with minimum degree m, for which $\langle k\rangle = 2m$. "Poisson" corresponds to $p(k) = \lambda e^{-\lambda}/k!$ with $\lambda = \langle k\rangle$, and

"Regular" stands for a random network for which all nodes have the same degree. In all cases, $N = 10\,000$.

V. AN ALGORITHM TO SWEEP THE RANGE OF POTENTIAL OVERLAPS BETWEEN TWO DEGREE SEQUENCES

In this section we design an algorithm that takes any pair of degree sequences D, D' and a value α between 0 and $\mathcal{O}^{\text{pot}}(D, D')$ and constructs a pair of networks G, G' with degree sequences D, D' whose overlap is as close as possible to α (values of α very close to $\mathcal{O}^{\text{pot}}(D, D')$ are not attainable since $\mathcal{O}^{\text{pot}}(D, D')$ is just an upper bound).

Assume that we have generated two random networks $G(0), G'(0)$ of N nodes using the configuration model. In view of (5), $\mathcal{O}(G(0), G'(0)) \approx 0$. Thus, it seems natural to propose an algorithm that works as follows. At each time step $t \geqslant 0$, modify the networks $G(t), G'(t)$ a little bit *without modifying the degree sequences* by performing a local operation (an operation involving few nodes and/or links) to obtain new networks $G(t+1), G'(t+1)$ in such a way that $\mathcal{O}(G(t+1), G'(t+1))$ is slightly larger than $\mathcal{O}(G(t), G'(t))$. Repeat until the overlap is close to α.

The kind of local operation that we will use in the scheme above is a *cross rewiring* [21], according to the following definition. Let $G(t), G'(t)$ be two networks of N nodes. A *good pair in $G(t)$ with respect to $G'(t)$* is a pair of links $\{a,b\}$, $\{c,d\}$ in $G(t)$ satisfying the following conditions:

(1) $\{a,b\}$ and $\{c,d\}$ are not links in $G'(t)$.

(2) $\{a,c\}$ and $\{b,d\}$ are not links in $G(t)$.

(3) $\{a,c\}$ is a link in $G'(t)$.

Analogously we define a *good pair in $G'(t)$ with respect to $G(t)$* by interchanging the roles of $G(t)$ and $G'(t)$ in the previous definition. Given a good pair $\{a,b\}$, $\{c,d\}$ in $G(t)$ with respect to $G'(t)$, the associated *cross-rewiring operation* consists of replacing the links $\{a,b\}$ and $\{c,d\}$ in $G(t)$ by $\{a,c\}$ and $\{b,d\}$ to get a new network $G(t+1)$. Observe that $G(t)$ and $G(t+1)$ are in general different as nonlabeled networks. However, the degrees of the involved nodes a,b,c,d are not modified after performing the cross rewiring. In consequence, $G(t)$ and $G(t+1)$ have the same degree sequences. On the other hand, set $G'(t+1) = G'(t)$ and let $E(t)$, $E(t+1)$, $E'(t)$, $E'(t+1)$ be respectively the sets of links of $G(t)$, $G(t+1)$, $G'(t)$, $G'(t+1)$. Then, $|E'(t+1)| = |E'(t)|$ and, by the definition of the cross-rewiring operation over a good pair, $|E(t+1)| = |E(t)|$. Moreover, by the definition of a good pair, either $|E(t+1) \cap E'(t+1)| = |E(t) \cap E'(t)| + 1$ if $\{b,d\}$ is a link in $G'(t)$ or $|E(t+1) \cap E'(t+1)| = |E(t) \cap E'(t)| + 2$ otherwise. Then, if we denote $\mathcal{O}(G(t), G'(t))$ and $\mathcal{O}(G(t+1), G'(t+1))$ by $\mathcal{O}(t)$ and $\mathcal{O}(t+1)$ respectively, a trivial computation yields that

$$\mathcal{O}(t+1) = \mathcal{O}(t) + \frac{x\mathcal{O}(t)^2 + 2x\mathcal{O}(t) + x}{L - x - x\mathcal{O}(t)}, \qquad (10)$$

where $x \in \{1,2\}$ and $L = |E(t)| + |E'(t)|$. As a consequence of (10), the overlap after performing a cross-rewiring operation in a good pair of links slightly (but strictly) increases.

From now on, let $0 \leqslant \alpha \leqslant \mathcal{O}^{\text{pot}}(D, D')$ be the desired overlap coefficient. In view of what has been said, let us consider the following *CR algorithm* (standing for "cross rewiring"):

CR algorithm [input: D, D', α]

(1) Use the configuration model to get two random networks $G(0), G'(0)$ of size N and degree sequences D, D'. The overlap between $G(0)$ and $G'(0)$ is close to 0.

At each time step $t \geqslant 0$:

(2) Choose at random (if it exists) a good pair of links in $G(t)$ with respect to $G'(t)$. Perform a cross-rewiring operation in $G(t)$ using such a pair, obtaining a new network $G(t+1)$. Set $G'(t+1) := G'(t)$. Then, by (10), $\mathcal{O}(G(t+1), G'(t+1)) > \mathcal{O}(G(t), G'(t))$. If $\mathcal{O}(G(t+1), G'(t+1)) \geqslant \alpha$, set $G := G(t+1)$, $G' := G'(t+1)$ and stop.

(3) Repeat the previous step interchanging the roles of $G(t)$ and $G'(t)$. Proceed to the next time step.

It is clear that after a finite number t_0 of steps the algorithm will stop, either because no good pairs are found or because the overlap between $G(t_0)$ and $G'(t_0)$ has reached the value α. In any case, the output of the algorithm is the pair of networks $G(t_0), G'(t_0)$. A natural question is whether in general the algorithm may halt forced by the condition that no good pairs are found, *before* having reached a value of the overlap close to α, especially when α is close to $\mathcal{O}^{\text{pot}}(D, D')$ [we stress the fact that $\mathcal{O}^{\text{pot}}(D, D')$ is just an upper bound, far from being realizable in general]. So, it makes sense to remove the stop condition given by the overlap and let the algorithm run until no more good pairs are found. In Table II we show the maximum overlap obtained in this way for several pairs of distributions, together with the upper bound $\mathcal{O}^{\text{pot}}(D, D')$. In all cases, the input degree sequences D, D' are random arrangements of the degree sets associated to the respective distributions. The obtained overlap is relatively close to the upper bound, suggesting that indeed the CR algorithm is able to sweep the entire range of *permitted* overlaps between 0 and $\mathcal{O}^{\text{pot}}(D, D')$.

TABLE II. Maximum overlap (first row) generated by the CR algorithm starting with two random arrangements D, D' of the corresponding degree sets vs the upper bound $\mathcal{O}^{\text{pot}}(D, D')$ (second row). In all cases $N = 10\,000$, $\langle k \rangle = 10$.

	Regular	Poisson	SF	Exponential
	1	0.738 11	0.562 01	0.637 31
Regular	1	0.776 12	0.611 29	0.678 17
		0.638 81	0.492 42	0.563 93
Poisson		0.696 05	0.565 36	0.625 55
			0.446 73	0.477 69
SF			0.514 43	0.538 22
				0.534 26
Exponential				0.589 36

VI. BELOW THE CRITICAL OVERLAP: TWO DESIRABLE STATISTICAL FEATURES OF THE NETWORKS GENERATED BY THE CR ALGORITHM

Let us introduce another relevant quantity that we will call *critical overlap*. It is defined as the potential overlap between two random sequences $(D_{\text{rand}}, D'_{\text{rand}}) \sim p(k) \times p'(k)$ where

$p(k), p'(k)$ are empirical distributions of N nodes:

$$\mathcal{O}^{cr}(N,p,p') := \mathcal{O}^{\text{pot}}(D_{\text{rand}}, D'_{\text{rand}}),$$

that for N big enough and pairs of distributions with bounded variance can be essentially considered as independent from the particular sampled sequences. Against an initial intuition, the critical overlap is not close to 0 but lies relatively close to $\overline{\mathcal{O}}$ (see Table I). By running the CR algorithm with sequences D_{rand}, D'_{rand} one can get any overlap α between 0 and (values close to) $\mathcal{O}^{cr}(N,p,p')$. As we will see, proceeding in this way the obtained two-layer network exhibits some desirable statistical features (lack of in- and cross-layer correlations). For higher values of α, it is unavoidable to introduce correlations and deviate from what happens in a "configuration model" context (Sec. VII).

A. Lack of in-layer degree-degree correlations

The lack of degree-degree correlations inside each layer is often a crucial requirement in the derivation of the equations governing mean-field multilayer models. In particular, this will be a basic assumption in the derivation of system (16) and (17) for the susceptible-infectious-recovered (SIR) model proposed in Sec. VIII. It is reasonable to expect that each network in a pair created via the CR algorithm with random initial sequences is uncorrelated, since:

(1) The networks $G(0), G'(0)$ are randomly generated via the configuration model algorithm, which is known to produce uncorrelated networks.

(2) A cross rewiring performed over a good pair of links $\{a,b\}$, $\{c,d\}$ increases (decreases) the global degree-degree correlation if the new links connect the two nodes with the smallest degrees and the two nodes with the largest degrees (respectively, if one of the new links connects the node with the largest degree to the node with lowest degree). But the rewiring criterion in the CR algorithm is intended to increase the overlap coefficient and has nothing to do with the degrees of the four involved nodes. So, some reconnections will increase the global degree-degree correlation and some will decrease it, thus expecting essentially an overall balance.

To support this claim, we show in Table III the standard Pearson coefficient r for each layer, computed from the two random variables defined by the degrees of the nodes at both ends of randomly chosen links [22]. Values of r close to -1 (respectively 1) account for dissortative (respectively assortative) networks, while values close to 0 correspond to uncorrelated networks. As in Table II, the CR algorithm was executed taking as input two random arrangements of the corresponding degree sets.

B. Lack of cross-layer degree-degree correlations

The *cross-layer degree-degree correlation* τ is defined as the correlation of the respective degrees k_i and k'_i of the same node i in the two layers. In Sec. VII we will show precisely how to measure it. We note that the epidemic model proposed in Sec. VIII will be simple enough to be independent of this sort of correlation, but this may not be the case for more sophisticated models, so that the question of obtaining a given overlap *controlling* τ makes sense. Observe that the cross-layer degree-degree correlation between two networks G, G' depends only on the respective degree sequences, not on the particular links joining the nodes in G and G'. On the other hand, $\tau \approx 0$ for two independent random sequences $D_{\text{rand}}, D'_{\text{rand}}$. Since during the execution of the CR algorithm the respective degree sequences are not modified, the lack of degree-degree correlations follows when using the CR algorithm starting with two independent random arrangements of the degree sets of $p(k), p'(k)$.

VII. ABOVE THE CRITICAL OVERLAP: ACCOUNTING FOR CROSS-LAYER DEGREE-DEGREE CORRELATIONS

In view of the previous sections, there is a natural algorithm that allows us to get any prescribed overlap $0 \leqslant \alpha \leqslant \overline{\mathcal{O}}(N,p,p')$: arrange the degree sets of $p(k)$ and $p'(k)$ to get degree sequences D, D' increasingly ordered. According to (9), $\overline{\mathcal{O}}(N,p,p') \leqslant \mathcal{O}^{\text{pot}}(D,D')$. Then, run the CR algorithm taking D, D' and α as input. This algorithm generates a pair of networks with *maximum* cross-layer degree-degree correlation. Indeed, nodes 1 and N have respectively the smallest and the largest degree in both layers, and the intermediate nodes have the same degree rank.

As we will see, there is an unavoidable relationship between high values of the overlap and the cross-layer degree-degree correlation, but the question arises whether it is possible to get a value of the overlap close to the maximum while controlling the cross-layer correlation to some extent.

Given two degree sequences $D = (k_1, k_2, \ldots, k_N)$ and $D' = (k'_1, k'_2, \ldots, k'_N)$, it is natural to measure the cross-layer degree-degree correlations by using the Kendall's τ-b coefficient [23]:

$$\tau(D,D') := \frac{N_c - N_d}{\sqrt{(N_0 - N_1)(N_0 - N_2)}}.$$

Here N_c is the number of concordant pairs, N_d is the number of discordant pairs, $N_0 = N(N-1)/2$, $N_1 = \sum_i t_i(t_i - 1)/2$, and $N_2 = \sum_j t'_j(t'_j - 1)/2$, where t_i is the number of tied values in the ith group of ties for D (analogously for t'_j and D'). A pair of indices $i \neq j$ is said to be *concordant* if $(k_i - k_j)(k'_i - k'_j) > 0$, *discordant* if $(k_i - k_j)(k'_i - k'_j) < 0$, or *tied* if $(k_i - k_j)(k'_i - k'_j) = 0$.

It is well known that if the agreement (respectively, disagreement) between the two rankings is perfect, then $\tau(D,D') = 1$ [respectively $\tau(D,D') = -1$], while if D and D' are independent (lack of cross-layer degree-degree correlation) then $\tau(D,D')$ is expected to be close to 0. Note also that if σ is any permutation, $\tau(\sigma(D),\sigma(D')) = \tau(D,D')$. So, in what follows we will assume without loss of generality that D is increasingly ordered:

$$k_1 \leqslant k_2 \leqslant \cdots \leqslant k_N.$$

The cross-layer degree-degree correlation between two networks G, G' depends only on the respective degree sequences D, D', not on the particular links joining the nodes in G and G'. Considering a permutation σ of the elements of D' corresponds to relabeling the nodes of G' to get a network G'' isomorphic (so, equally distributed) to G', and it makes sense to study how the potential overlap $\mathcal{O}^{\text{pot}}(D,\sigma(D'))$ and the correlation coefficient $\tau(D,\sigma(D'))$ vary in terms of σ with respect to $\mathcal{O}^{\text{pot}}(D,D')$ and $\tau(D,D')$. Since any permutation decomposes

TABLE III. Pearson coefficient to measure the degree-degree correlations in each layer for several pairs of networks obtained from the CR algorithm with prescribed overlap $\alpha = 0.15, 0.3, 0.45$. In all cases, $N = 10\,000$ and $\langle k \rangle = 10$.

	$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.45$		$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.45$
Poisson	0.022 88	0.024 74	0.054 29	Poisson	0.014 04	0.037 58	0.073 38
SF	0.006 73	0.047 74	0.129 42	Poisson	0.013 82	0.043 92	0.056 24
SF	0.004 19	0.032 07	0.074 98	SF	0.008 30	0.030 33	0.078 82
Exponential	0.027 11	0.077 71	0.131 75	SF	0.015 86	0.047 19	0.079 09
Poisson	0.018 88	0.035 88	0.054 01	Exponential	0.022 10	0.070 99	0.128 41
Exponential	0.032 90	0.070 54	0.097 41	Exponential	0.052 03	0.078 87	0.117 22

in a sequence of transpositions (or swaps) of two elements, let us consider a pair of indices $i < j$ such that $k'_i > k'_j$ (a discordant pair). When we swap both entries in D' to get a sequence D'' such that $k''_i = k'_j$, $k''_j = k'_i$ and $k''_l = k'_l$ for $l \neq i, j$, then $\tau(D, D'') > \tau(D, D')$. On the other hand, it is trivial to check that $\sum_n \min\{k_n, k''_n\} - \sum_n \min\{k_n, k'_n\}$ equals

(a) 0 if $k'_j < k'_i < k_i < k_j$,
(b) $k_j - k_i > 0$ if $k'_j < k_i < k_j < k'_i$,
(c) $k'_i - k_i > 0$ if $k'_j < k_i < k'_i < k_j$,
(d) $k'_i - k'_j > 0$ if $k_i < k'_j < k'_i < k_j$,
(e) $k_j - k'_j > 0$ if $k_i < k'_j < k_j < k'_i$,
(f) 0 if $k_i < k_j < k'_j < k'_i$.

Since $\mathcal{O}^{\text{pot}}((r_n)_1^N, (r'_n)_1^N)$ is increasing as a function of $\sum_n \min\{r_n, r'_n\}$ [see (7)], it follows that the potential overlap does not decrease when performing a swap that increases the τ-b coefficient. Analogously, one can check that the τ-b coefficient does not decrease after a swap that increases the potential overlap. This remark plainly shows that, as expected, there is a direct relationship between overlap and cross-layer degree-degree correlation.

Keeping in mind that we want to find a sequence of swaps in order to increase the potential overlap while controlling in some sense the cross-layer degree-degree correlation, a crucial remark is that, together with the swaps of types (b)–(e) above, that increase both the potential overlap and the τ-b coefficient, there are two cases for which the swap $k'_i \leftrightarrow k'_j$ does not modify the potential overlap while it *decreases* the τ-b coefficient:

(A) $k'_i < k'_j < k_1 < k_j$,
(B) $k_i < k_j < k'_i < k'_j$.

Before describing what we call the LS-CR algorithm (standing for *label swap–cross rewiring*), we give an example of how it works. Let $p(k), p'(k)$ be two empirical distributions approximating respectively a Poisson distribution with $\langle k \rangle = 10$ and a scale-free distribution with $\langle k' \rangle = 12$. Set $N = 10\,000$. Let D, D' two random arrangements of the degree sets. The cross-layer degree-degree correlation is expected to be close to 0. Indeed, in a particular simulation we get $\tau(D, D') = 0.014\,70$, while $\mathcal{O}^{\text{pot}}(D, D') = 0.574\,78 = \mathcal{O}^{cr}(N, p, p')$. So, since the cross-rewiring operations do not modify the cross-layer correlation, if we want a prescribed overlap α smaller than 0.574 78, the CR algorithm suffices to construct a two-layer network with overlap close to α and a small τ-b coefficient. But suppose that the desired overlap is significantly larger. To see how big it can be, rearrange the elements in D, D' to get two sequences $\sigma(D), \rho(D')$ increasingly ordered and compute $\mathcal{O}^{\text{pot}}(\sigma(D), \rho(D')) = 0.749\,283$, that according to (9) is an absolute upper bound for the largest permitted overlap. The corresponding τ-b coefficient is of course very close to 1: $\tau(\sigma(D), \rho(D')) = 0.949\,190$. Suppose now that the desired overlap is very close to $\overline{\mathcal{O}}$, for instance $\alpha = 0.73$. We proceed as follows. Rearrange D' using the permutation σ [so $\sigma(D)$ is increasingly ordered while $\sigma(D')$ is *not*]. Both the potential overlap and the τ-b coefficient between $\sigma(D)$ and $\sigma(D')$ do not change. Now we perform a series of swaps in $\sigma(D')$ of any of types (b)–(e), that increase both the potential overlap and the τ-b coefficient, until we reach the potential overlap α. Then, we perform as many swaps of type (A)-(B) as possible in order to diminish the τ-b coefficient without modifying the potential overlap. After running this algorithm in our particular simulation, we get a sequence D'' such that $\tau(\sigma(D), D'') = 0.648\,549$. Of course the correlation is high, but significantly smaller than 1. Finally, now we can use the CR algorithm with input $\sigma(D), D'', \alpha$ to effectively construct the two-layer network. If we repeat the previous scheme with a prescribed overlap $\alpha = 0.65$, still close to the maximum, we get a sequence D'' such that $\tau(\sigma(D), D'') = 0.209\,785$. It is instructive to visualize the evolution of both $\mathcal{O}^{\text{pot}}$ and τ during the complete sequence of swaps (see Fig. 1).

So, let $0 \leqslant \alpha \leqslant \overline{\mathcal{O}}(N, p, p')$ be the desired overlap. The following *LS-CR algorithm* (standing for *label swap–cross rewiring*) is intended to construct two networks of N nodes distributed according to $p(k), p'(k)$ with an overlap close to α and a cross-layer degree-degree correlation as small as possible.

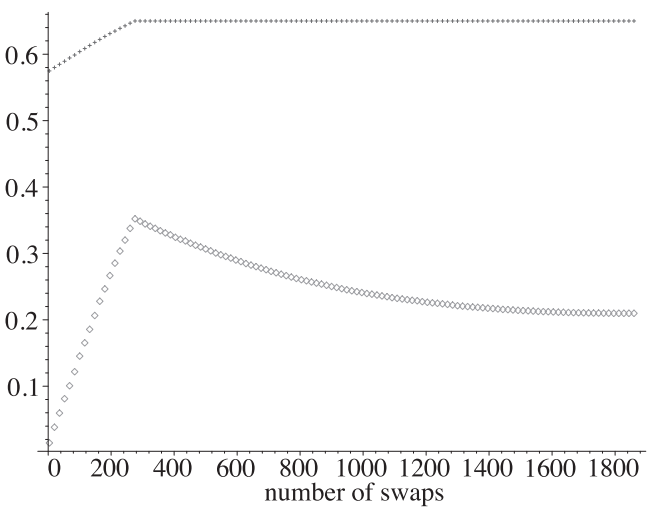

FIG. 1. Evolution of the potential overlap (crosses) and the cross-layer degree-degree correlation (diamonds) when performing a sequence of swaps.

TABLE IV. Three examples of a series of executions of the LS-CR algorithm with prescribed overlaps 0.6, 0.65, 0.7, 0.75, 0.8, and 0.85. In all cases, $N = 10\,000$, $\langle k\rangle = 12$ for the first distribution, and $\langle k'\rangle = 14$ for the second one. For any pair of distributions we report both the critical and the theoretical maximum overlap. For each two-layer network, we show the overlap α, the Kendall's τ-b coefficient τ for the cross-layer degree-degree correlation, and the Pearson coefficients ρ_1, ρ_2 for the in-layer degree-degree correlations.

		0.6	0.65	0.70	0.75	0.80	0.85
	α	0.5844	0.5952	0.6389	0.6985	0.7614	0.8050
ER 12–Exp 14	τ	−0.0178	−0.0005	0.0935	0.3537	0.6152	0.7899
$\mathcal{O}^{cr} = 0.6330$	ρ_1	0.0753	0.0692	0.0522	0.0241	−0.0087	−0.0095
$\overline{\mathcal{O}} = 0.8350$	ρ_2	0.1865	0.2055	0.3517	0.3543	0.3311	0.2727
	α	0.5387	0.5788	0.6386	0.7014	0.7688	0.8459
Exp 12–Exp 14	τ	−0.0006	0.0323	0.2116	0.3656	0.5132	0.6743
$\mathcal{O}^{cr} = 0.5797$	ρ_1	0.2010	0.1833	0.1451	0.0851	0.0366	0.0166
$\overline{\mathcal{O}} = 0.8545$	ρ_2	0.1715	0.2377	0.2104	0.1798	0.3271	0.1102
	α	0.5102	0.5702	0.6236	0.6879	0.7611	0.8385
SF 12–SF 14	τ	0.1543	0.2900	0.4033	0.4989	0.5722	0.6246
$\mathcal{O}^{cr} = 0.5027$	ρ_1	0.0982	0.0957	0.0761	0.0551	0.0269	0.0021
$\overline{\mathcal{O}} = 0.8522$	ρ_2	0.1215	0.1106	0.0980	0.0696	0.0362	−0.0760

LS-CR algorithm [input: $N, p(k), p'(k), \alpha$]

(1) Take degree sequences $D_{\text{rand}}, D'_{\text{rand}}$ by rearranging at random the degree sets of $p(k), p'(k)$. Then,

$$\mathcal{O}^{\text{pot}}(D_{\text{rand}}, D'_{\text{rand}}) = \mathcal{O}^{cr}(N, p, p').$$

(2) If $\alpha \leqslant \mathcal{O}^{cr}(N, p, p')$, execute the CR algorithm with input $D_{\text{rand}}, D'_{\text{rand}}, \alpha$ and stop. Otherwise,

(3) Let σ be the permutation that rearranges D_{rand} increasingly. Set $D_0 = \sigma(D_{\text{rand}})$, $D'_0 = \sigma(D'_{\text{rand}})$. Then,

$$\mathcal{O}^{\text{pot}}(D_0, D'_0) = \mathcal{O}^{\text{pot}}(D_{\text{rand}}, D'_{\text{rand}}),$$
$$\tau(D_0, D'_0) = \tau(D_{\text{rand}}, D'_{\text{rand}}) \approx 0.$$

(4) At each time step $t \geqslant 0$:

Choose at random (if it exists) a pair of indices $i < j$ such that the four corresponding entries in D_0 and D'_t satisfy any of the conditions (b)–(e). Swap the entries i and j in D'_t to get a new sequence D'_{t+1}. Then,

$$\mathcal{O}^{\text{pot}}(D_0, D'_{t+1}) > \mathcal{O}^{\text{pot}}(D_0, D'_t).$$

If $\mathcal{O}^{\text{pot}}(D_0, D'_{t+1}) \geqslant \alpha$, set $t_0 := t + 1$ and go to step 5. Otherwise, proceed to the next time step.

(5) At each time step $t \geqslant t_0$:

Choose at random (if it exists) a pair of indices $i < j$ such that the four corresponding entries in D_0 and D'_t satisfy either (A) or (B). Swap the entries i and j in D'_t to get a new sequence D'_{t+1}. Then,

$$\mathcal{O}^{\text{pot}}(D_0, D'_{t+1}) = \mathcal{O}^{\text{pot}}(D_0, D'_t),$$
$$\tau(D_0, D'_{t+1}) < \tau(D_0, D'_t).$$

If no pairs are found satisfying (A) or (B), set $t_1 := t$ and go to step 6. Otherwise, proceed to the next time step.

(6) Execute the CR algorithm with input D_0, D'_{t_1}, α.

In Table IV we show some statistical features of the two-layer network obtained from the LS-CR algorithm for several pairs of distributions and different values of the prescribed overlap, all beyond the critical one. In each case we show the obtained overlap α, the Kendall's τ-b coefficient τ for the cross-layer correlation, and the Pearson coefficients ρ_1, ρ_2 for the degree-degree correlation inside each layer. The evolution of the statistics with the overlap depends of course on the particular distributions considered, but some clear general conclusions can be extracted. In all cases, the obtained overlaps are close to the prescribed one. The τ-b coefficient approaches 1 (even relatively) only for values of the overlap beyond about 80% of the theoretical maximum. The degree-degree correlations inside each layer remain in most cases close to 0.

As a final remark, it is clear that the LS-CR algorithm admits a lot of variants depending on the type and order of swaps that one performs (in the "LS" part of the algorithm). For instance, one may be interested in inverting the roles and generate a two-layer network with a prescribed cross-layer degree-degree correlation, while getting an overlap as big as possible.

VIII. A SIMPLE EXAMPLE: A MEAN-FIELD SIR EPIDEMIC MODEL ON A TWO-LAYER NETWORK

This section aims at illustrating that, specially for mean-field models of processes that take place over a two-layer network, the qualitative response of a model may depend critically on the interlayer overlap. To do it, we present a simple example of an epidemic model with information dissemination and determine the impact of the overlap on the final outbreak size predicted by the model.

Epidemic models describe the spread of infectious diseases on populations whose individuals are classified into distinct classes according to their infection state as, for instance, susceptible (S), infectious (I), and recovered (R) individuals. A closer look at the physical transmission of an infection reveals that a suitable description of populations must take into account the network layer A of physical contacts among individuals, with nodes representing individuals and links corresponding to physical contacts along which disease can propagate. On the other hand, if one assumes that the probability of getting infected through an infectious contact S-I depends on the awareness state of the susceptible individual, then a second network layer B over which information about the infection status of individuals circulates can be considered. In the context of management and control of sexually transmitted diseases

(STDs), an example of this second network layer is given by the partner notification program. This service helps to reach sexual contacts of patients of STDs and inform them that they may be at risk, and hence the need of seeking medical care [24,25]. So, in our approach, if a pair of individuals, one susceptible and the other infectious, are connected to each other in both network layers, we assume that the transmission rate β_c (here c stands for *common*) will be smaller than the normal transmission rate β because the susceptible partner adopts preventive measures to diminish the risk of contagion.

According to this scenario, next we derive a mean-field SIR epidemic model which implicitly assumes spreading of information on the infection status of nodes in one layer, while explicitly modeling the transmission of an infectious agent in a second layer. Following the standard approach for STDs where the heterogeneity in the number of contacts (sexual partners) is a basic ingredient [26], individuals are classified according to their infection state and their number of physical contacts. So, the model will take into account the network layer A of physical contacts in terms of its degree distribution $p_A(k) = N_k/N$ where N_k is the number of individuals having degree k. Analogously, the information or notification network (network layer B) is described by its degree distribution $p_B(k)$. For the sake of brevity, a pair of nodes connected to each other in both networks is said to share a *common link*, although the natures of the connections are dissimilar. Moreover, the model does not assume that links in layer B are a subset of those in layer A, as could be the case in partner notification.

Within each layer, it is assumed that there is no degree-degree correlation, i.e., neighbors in each layer are randomly sampled from the population according to the so-called proportionate mixing of individuals. This means that, in each layer, the probability $P(k'|k)$ that a node of degree k is connected to a node of degree k' is independent of the degree k and it is given by the fraction of links pointing to nodes of degree k', i.e., $P(k'|k) = k'p(k')/\langle k\rangle$. Now, let $I_k(t)$ be the number of infectious nodes of degree k at time t in layer A. Although links are unordered pairs of connected nodes by definition, let us consider that every link $\{u,v\}$ gives rise to two *oriented links* $u \to v$ and $v \to u$. Then, the probability that a randomly chosen oriented link of A leads to an infectious node is given by the fraction of oriented links in A pointing to infectious nodes [26], that is,

$$\Theta_I(t) = \frac{1}{\langle k_A\rangle N}\sum_k k\, I_k(t) = \frac{1}{\langle k_A\rangle}\sum_k k\, i_k(t),$$

where $\langle k_A\rangle$ is the average degree in A, and $i_k(t) := I_k(t)/N$ is the fraction of nodes that are both infectious and of degree k in A at time t.

Finally, let L_A, L_B, and $L_{A\cap B}$ denote the number of links of A, B, and common links, respectively. Let $p_{B|A}$ be the probability that a randomly chosen link of A, an A link, connects two nodes that are also connected in B, that is, $p_{B|A} = \frac{L_{A\cap B}}{L_A}$. Similarly, $p_{A|B} = \frac{L_{A\cap B}}{L_B}$ is the probability that a randomly chosen B link is a common link to both networks.

We stress that a key assumption in the model derivation is the uniformity of the overlap between the links of each layer. More precisely: the overlap α is a global feature of the pair of networks $\{A,B\}$ that depends on the respective whole sets of links, and the equations of the model, that will account for what happens around a typical node i, will be derived using α as a parameter. Implicitly, this corresponds to the mean-field approximation that the *local* overlap around the node i (fraction of links confluent to i in the union network that are common links of A and B) does not deviate significantly from α. This assumption is clearly unrealistic in general. For instance, a particular run of the LS-CR algorithm with prescribed overlap $\alpha = 0.5$ over two exponential networks of 5000 nodes and mean degrees 45 and 30 leads to a mean local overlap equal to 0.5439 and a standard deviation of 0.1790. So, it is relevant to test the goodness of this approximation by comparing the predictions of the model with simulation outputs.

Taking it all into account, the epidemic spreading is described in terms of $I_k(t)$, and also of $S_k(t)$ and $R_k(t)$, the number of susceptible and recovered nodes of degree k in layer A at time t respectively, which satisfy $S_k(t) + I_k(t) + R_k(t) = N_k$. In particular, the differential equations for S_k and I_k are

$$\frac{dS_k}{dt} = -k(1-p_{B|A})\beta S_k\,\Theta_I - k p_{B|A}\,\beta_c S_k \Theta_I, \tag{11}$$

$$\frac{dI_k}{dt} = k(1-p_{B|A})\beta S_k\,\Theta_I + k p_{B|A}\,\beta_c S_k \Theta_I - \mu I_k. \tag{12}$$

The first term on the right-hand side of (12) is the rate of creation of new infectious nodes of degree k in A due to transmissions of the infection through links that only belong to layer A, whereas the second one is the rate of creation of new infectious nodes from transmissions across common links. The last term accounts for the recoveries of infectious nodes, which occur at a recovery rate μ. Here $\langle k_A\rangle p_{B|A}$ is the expected number of common oriented links. Therefore, since this number is the same regardless of the network we use to compute it, the following consistency relationship must follow:

$$\langle k_A\rangle p_{B|A} = \langle k_B\rangle p_{A|B}. \tag{13}$$

Now let us express $p_{B|A}$ and $p_{A|B}$ in terms of the overlap $\alpha := \mathcal{O}(A,B)$, which is defined as $\alpha = \frac{L_{A\cap B}}{L_{A\cup B}}$ where $L_{A\cup B}$ is the set of links of the union network $A \cup B$. Using that $\langle k\rangle N = 2L$, $p_{B|A}$ can be expressed in terms of α as follows:

$$\begin{aligned} p_{B|A} &= \frac{L_{A\cap B}}{L_A} = \frac{L_{A\cap B}}{L_{A\cup B}}\frac{L_{A\cup B}}{L_A} \\ &= \alpha\,\frac{L_A + L_B - L_{A\cap B}}{L_A} = \alpha\left(1 + \frac{\langle k_B\rangle}{\langle k_A\rangle} - p_{B|A}\right). \end{aligned} \tag{14}$$

From this simple relationship it immediately follows that

$$p_{B|A} = \left(1 + \frac{\langle k_B\rangle}{\langle k_A\rangle}\right)\frac{\alpha}{1+\alpha}. \tag{15}$$

Similarly,

$$p_{A|B} = \left(1 + \frac{\langle k_A\rangle}{\langle k_B\rangle}\right)\frac{\alpha}{1+\alpha}.$$

As expected, $p_{B|A}$ and $p_{A|B}$ fulfill relationship (13).

Introducing (15) into system (11) and (12), the overlap appears as a new parameter of the model which now, in terms of the fractions $s_k = S_k/N$ and $i_k = I_k/N$ of susceptible and

infectious nodes of degree k, reads

$$\frac{ds_k}{dt} = -k\beta_0(\alpha)s_k\,\Theta_I, \tag{16}$$

$$\frac{di_k}{dt} = k\beta_0(\alpha)s_k\,\Theta_I - \mu i_k, \tag{17}$$

where

$$\beta_0(\alpha) := \frac{1}{1+\alpha}\left[\beta\left(1-\frac{\langle k_B\rangle}{\langle k_A\rangle}\alpha\right)+\beta_c\left(1+\frac{\langle k_B\rangle}{\langle k_A\rangle}\right)\alpha\right],$$

and $s_k + i_k + r_k = p_A(k)$.

These equations correspond to the standard SIR model for heterogeneous and closed populations with proportionate mixing [26,27], but with an averaged transmission rate $\beta_0(\alpha)$ that takes into account the degree of overlap between the two layers. A similar mean-field approach for modeling epidemic spreading in single heterogeneous networks was adopted in [28] using, as a state variable, the fraction ρ_k of nodes of degree k that are infectious. The connection between both approaches is given by the relationship between the state variables. For instance, $i_k = I_k/N = I_k/N_k \cdot N_k/N =: \rho_k\, p(k)$.

Simple facts about system (16) and (17) are as follows:

(1) Since the factor $\alpha/(1+\alpha)$ in (15) is increasing in α, and $\alpha \leqslant \min\{\langle k_A\rangle,\langle k_B\rangle\}/\max\{\langle k_A\rangle,\langle k_B\rangle\}$ [see (8)], it follows that

$$p_{B|A} \leqslant \frac{\min\{\langle k_A\rangle,\langle k_B\rangle\}}{\langle k_A\rangle}.$$

So, when $\langle k_A\rangle \leqslant \langle k_B\rangle$ we get $p_{B|A} \leqslant 1$ while for $\langle k_A\rangle > \langle k_B\rangle$ we get $p_{B|A} \leqslant \langle k_B\rangle/\langle k_A\rangle < 1$.

(2) If $\beta_c = \beta$ or $\alpha = 0$, the system reduces to the classic SIR model, as expected, because information dissemination plays no role in the infection spread. If $\alpha = 1$, we actually have one network and again the system reduces to the SIR model but now with β replaced by β_c.

To determine the impact of the network overlap on the initial epidemic growth, we linearize the system (16) and (17) about the disease-free equilibrium $(s_k^*,i_k^*) = (p_A(k),0)\,\forall k$ and obtain that the elements of the Jacobian matrix J^* evaluated at this equilibrium are

$$J^*_{kk'} = \frac{\beta_0(\alpha)}{\langle k_A\rangle}kk'p_A(k) - \mu\delta_{kk'},$$

where $\delta_{kk'}$ is the Kronecker delta. Since the only non-zero eigenvalue of the matrix $(kk'p_A(k))$ is equal to $\langle k_A^2\rangle = \sum_k k^2 p_A(k)$ [with an associated eigenvector whose components v_k are proportional to $kp_A(k)$], it follows that the largest eigenvalue of J^* is

$$\Lambda_1(\alpha) = \frac{\langle k_A^2\rangle}{\langle k_A\rangle}\beta_0(\alpha) - \mu,$$

which corresponds to the initial growth rate of the epidemic. Clearly, Λ_1 decreases with α because $\beta_c < \beta$, and $\Lambda_1(\alpha) = 0$ at $\beta_0(\alpha)/\mu = \langle k_A\rangle/\langle k_A^2\rangle$, which corresponds to the epidemic threshold according to this mean-field approximation. Notice that, under proportionate mixing, the expected degree of a node reached by following a randomly chosen link in network A is $\langle k_A^2\rangle/\langle k_A\rangle$.

We have checked the accuracy of the model (16) and (17) by collating the predicted epidemic final size, i.e., the number of individuals ever infected, with the histogram of final outbreak sizes of an ensemble of 1500 stochastic epidemic realizations on a network of 5000 nodes and using α as a tuning parameter. Each network layer is generated according to the configuration model, and the desired value of α is attained using the CR algorithm to guarantee the in-layer degree-degree correlation is as close to 0 as possible. To clearly separate the dichotomy "minor outbreak vs major outbreak" (initial extinctions are highly feasible because only one node is randomly infected at $t = 0$), we chose the values of the parameters to be far enough from the epidemic threshold. This guarantees the existence of a marked distribution of final major-outbreak sizes, in addition to the one of minor-outbreak sizes around 1.

For an acceptable prediction of the model, the final epidemic size obtained from the mean-field approximation should be relatively close to the mean value around which major outbreaks are distributed. We insist that, in addition to the well-known limitations of the mean-field approach when modeling epidemic processes on one-layer networks [29], here the accuracy of predictions also depends on the fulfillment of the implicit hypotheses assumed in the derivation of expression (15) for $p_{B|A}$. Namely, (i) there is no in-layer degree-degree correlations, and (ii) the occurrence of a common link is the same for any pair of nodes in the network. So, the value of $p_{B|A}$ does not depend on the degree of a node in layer A and, hence, the overlap between layers is uniformly distributed (i.e., there are no parts of the network more overlapped than others). Assumption (i) is guaranteed by the algorithms. However, assumption (ii) is not feasible when the architectures of both layers are very different from each other. Then, our simulations have been performed on networks with two different architectures reflecting two extreme cases. First, we have considered two-layer networks where each layer is in turn a regular random network. This guarantees that both hypotheses are satisfied and, moreover, a good accuracy of the mean-field approach for this type of network if the degree of each layer is high enough. Second, we have considered networks with both layers having exponential degree distributions which have a high variance. In both cases, the mean degrees are 45 (layer A) and 30 (layer B), both high enough to minimize the impact of stochastic fluctuations around infected nodes.

To derive an analytical expression of the final epidemic size note that, for all k, $S_k(\infty) = N_k - R_k(\infty)$ since $\lim_{t\to\infty} I_k(t) = 0$ [$S_k(\infty)$ and $R_k(\infty)$ are the limits of $S_k(t)$ and $R_k(t)$ as $t \to \infty$]. From this fact, the initial condition is $(s_k(0),i_k(0)) = (p_A(k),0)$, and integrating from 0 to ∞ the equation resulting from the sum of Eqs. (16) and (17), we have

$$\int_0^\infty I_k(t)dt = \frac{1}{\mu}R_k(\infty).$$

Now, integrating (16) from 0 to ∞, and using the previous expression, it follows that $R_k(\infty) = N_k(1-e^{-k\xi})$, with $\xi := \frac{\beta_0(\alpha)}{\mu\langle k_A\rangle N}\sum_k kR_k(\infty)$. Therefore, the final epidemic size is given by

$$\sum_k R_k(\infty) = \sum_k N_k(1-e^{-k\xi}) \tag{18}$$

with ξ being the positive solution (if it exists) of the equation

$$\xi = \frac{\beta_0(\alpha)}{\langle k_A\rangle\mu}\langle k(1-e^{-k\xi})\rangle. \tag{19}$$

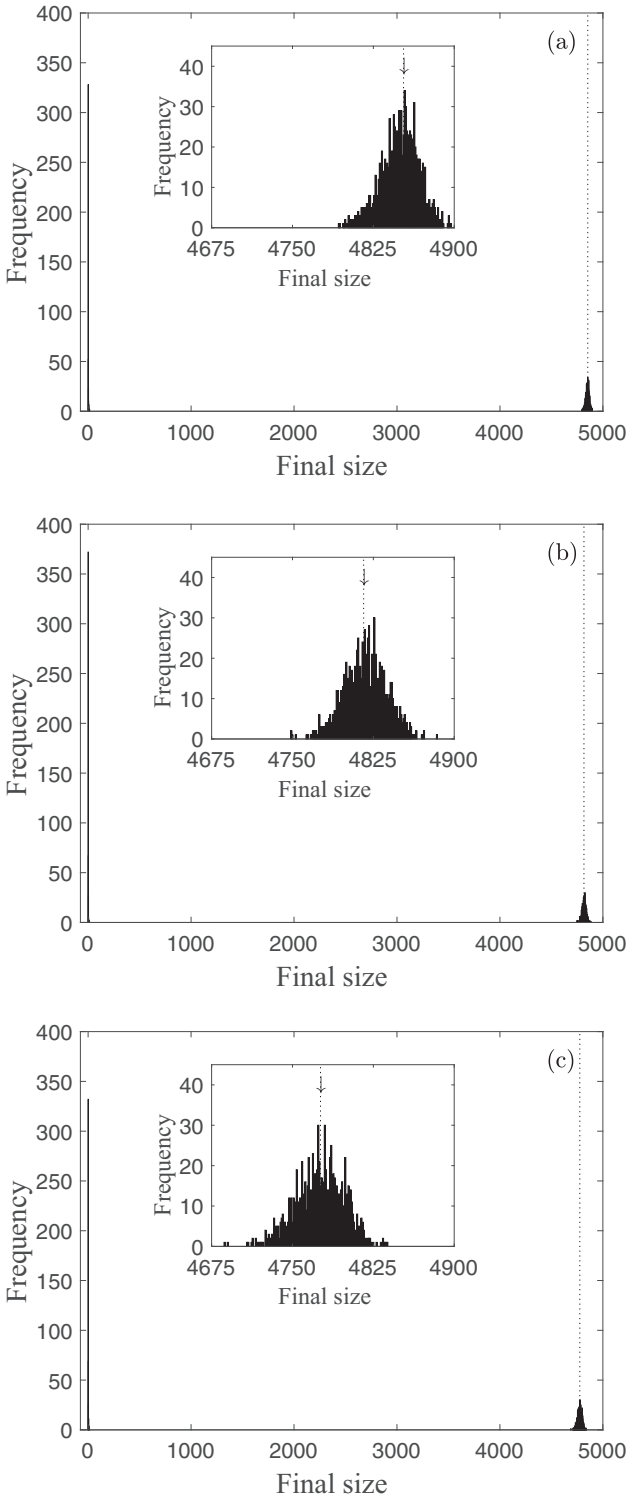

FIG. 2. Histograms of 1500 final outbreak sizes on a two-layer network of 5000 nodes. Each layer is generated as a regular random network of degree $k_A = 45$ and $k_B = 30$, respectively. The size distribution of small outbreaks ranges from 1 to 9 in the three panels but only the frequency of a final size equal to 1 (the initial infected node recovers before infecting any neighbor) can be distinguished. Vertical dotted line from bottom to top shows the predicted final epidemic size according to (18) and (19). Insets: magnified histograms of major-outbreak sizes. Parameters: $\beta = 0.1$, $\beta_c = 0.05$, $\mu = 1$, and $\alpha = 0.3$ (a), 0.4 (b), and 0.5 (c).

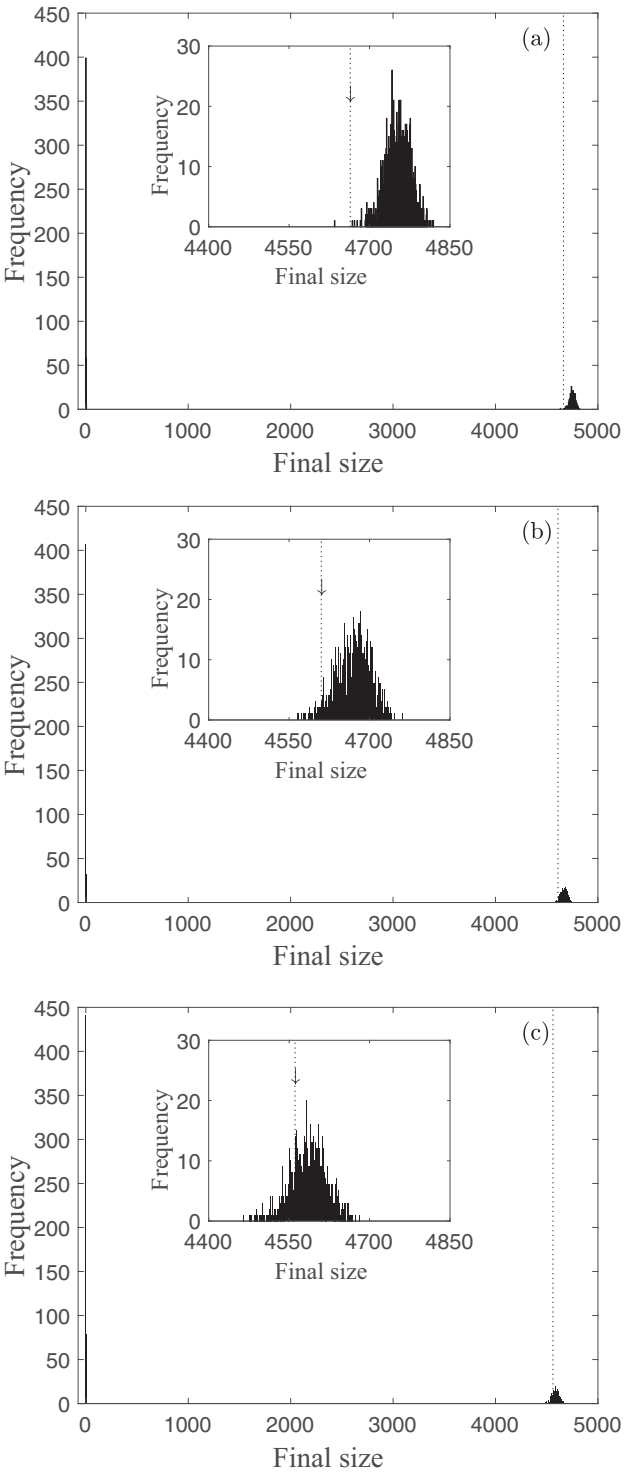

FIG. 3. Histograms of 1500 final outbreak sizes on a two-layer network of 5000 nodes and exponential degree distributions on each layer with expected degrees $\langle k_A \rangle = 45$ and $\langle k_B \rangle = 30$, respectively (see Sec. IV for details). The size distribution of small outbreaks ranges from 1 to 5 in top panel, and from 1 to 7 in middle and bottom panels. Note that only the frequency of a final size equal to 1 (the initial infected node recovers before infecting any neighbor) can be perceived. Vertical dotted line from bottom to top shows the predicted final epidemic size according to (18) and (19). Insets: magnified histograms of major-outbreak sizes. Parameters: $\beta = 0.1$, $\beta_c = 0.05$, $\mu = 1$, and $\alpha = 0.3$ (a), 0.4 (b), and 0.5 (c).

It is interesting to observe that, since $\xi = 0$ is always a solution, this equation will have a unique positive solution $\xi^*(\alpha)$ if the derivative with respect to ξ of its right-hand side is larger than 1, which is equivalent to having $\Lambda_1(\alpha) > 0$, i.e., a positive initial epidemic growth (the uniqueness follows from the convexity of the function defined by this right-hand side). This derivation of the final epidemic size is presented for the sake of completeness because, indeed, it follows from the one given in Ref. [26], Appendix E, in a more general setting.

Figures 2 and 3 show the histograms of final outbreak sizes obtained from the stochastic simulations on regular random networks (RNNs) and on networks with exponential degree distributions, respectively, for three values of α. These figures also show the mean final epidemic size given by Eq. (18) after numerically solving Eq. (19) using the generated degree sequence in layer A. When the CR algorithm is applied and the distribution of major outbreaks is clearly distinguished from the one of minor outbreaks (otherwise to talk about the final size of an epidemic makes no sense), the predicted final size on RNNs is almost the same as the mean final size of the major outbreaks (4853 vs 4850, 4816 vs 4818, and 4776 vs 4775 for $\alpha = 0.3$, 0.4, 0.5, and rounded values). For the exponential networks, the predicted final epidemic size differs less than 5% from the mean final size of major outbreaks for most of the networks generated with the configuration model with different degree distributions (not shown here). In Fig. 3, this disagreement is indeed less than 2%. Therefore, in these cases the proposed mean-field model qualitatively captures the impact of the overlap on the expected final size of an epidemic.

IX. CONCLUSIONS

The aim of this paper is to provide a toolbox of algorithms to generate two networks G, G' sharing the same set of N nodes (a two-layer network) whose respective degree distributions $p(k), p'(k)$ are given, *with a prescribed overlap coefficient* α defined by the Jaccard index.

First of all, we study the possible range $(\alpha_m, \alpha_M) \subset [0,1]$ of permitted overlap coefficients in terms of $p(k)$, $p'(k)$ and N. We start by proving that $\alpha_m \approx 0$ for any $p(k), p'(k)$ and N big enough. Given two *fixed* degree sequences $D = (k_i)_{i=1}^N$ and $D' = (k'_i)_{i=1}^N$, we derive an upper bound of α for any pair of networks sequenced as D and D', by assuming the condition (not realizable in general) that the intersection network has exactly $\min\{k_i, k'_i\}$ links attached at node i. We call this upper bound *potential overlap* between D and D' and we denote it by $\mathcal{O}^{\text{pot}}(D, D')$. Then we prove that an estimate (more properly, an upper bound) for α_M is precisely $\mathcal{O}^{\text{pot}}(S, S')$, where S, S' are degree sequences sampled from $p(k), p'(k)$ whose respective elements are increasingly arranged.

To construct the desired algorithm we proceed in three steps. First, we define a partial procedure, that we call *CR algorithm*, that takes any pair of degree sequences D, D' and a value α between 0 and $\mathcal{O}^{\text{pot}}(D, D')$ and constructs a pair of networks G, G' sequenced as D, D' whose overlap is as close as possible to α. Second, we introduce what we call the *critical overlap* $\alpha_m < \alpha_c < \alpha_M$, defined as $\mathcal{O}^{\text{pot}}(D_{\text{rand}}, D'_{\text{rand}})$ where $D_{\text{rand}}, D'_{\text{rand}}$ are random sequences sampled from $p(k), p'(k)$. Against an initial intuition, α_c is closer to α_M than expected. We show that the CR algorithm with $D_{\text{rand}}, D'_{\text{rand}}$ as input suffices to construct a pair of networks having any overlap below α_c and exhibiting some desirable statistical properties, specifically lack of in- and cross-layer degree-degree correlations. Finally, when the desired overlap is beyond α_c, we propose what we call the *LS-CR algorithm*, that minimizes the cross-layer degree-degree correlations that unavoidably appear for high values of α.

To illustrate the impact of the network overlap, we present a simple example of an SIR epidemic model over a two-layer network (physical contacts–information dissemination) and determine the impact of α on the initial epidemic growth and on the final epidemic size predicted by the model. The comparison of the epidemic final size with the average final size of major outbreaks obtained from stochastic simulations shows an excellent agreement on regular random networks of high degrees, and a qualitatively good agreement on exponential networks. Provided that one wants to perform simulations to validate this (or any) model where a multilayer network is involved, a systematic procedure to generate couples of networks of given size and degree distributions with a prescribed value of the overlap α as those presented here seems to be a useful tool.

ACKNOWLEDGMENTS

The authors have been partially supported by Research Grants No. MTM2014-52402-C3-3-P (J.S., D.J.), No. MTM2017-86795-C3-1-P (D.J.) of Ministerio de Economía y Competitividad (MINECO) of the Spanish government, No. 2017SGR1617 (D.J.), No. 2014SGR1083 (J.S.) of the Generalitat de Catalunya, and No. MPCUdG2016/047 of the Universitat de Girona.

[1] F. D. Sahneh and C. Scoglio, Competitive epidemic spreading over arbitrary multilayer networks, Phys. Rev. E **89**, 062817 (2014).

[2] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, Mathematical Formulation of Multilayer Networks, Phys. Rev. X **3**, 041022 (2013).

[3] G. D'Agostino and A. Scala, *Networks of Networks: The Last Frontier of Complexity* (Springer, New York, 2014).

[4] A. Saumell-Mendiola, M. A. Serrano, and M. Boguná, Epidemic spreading on interconnected networks, Phys. Rev. E **86**, 026106 (2012).

[5] S. Funk and V. A. A. Jansen, Interacting epidemics on overlay networks, Phys. Rev. E **81**, 036118 (2010).

[6] C. Granell, S. Gómez, and A. Arenas, Dynamical Interplay Between Awareness and Epidemic Spreading in Multiplex Networks, Phys. Rev. Lett. **111**, 128701 (2013).

[7] C. Granell, S. Gómez, and A. Arenas, Competing spreading processes on multiplex networks: awareness and epidemics, Phys. Rev. E **90**, 012808 (2014).

[8] X. Wei, N. C. Valler, B. A. Prakash, I. Neamtiu, M. Faloutsos, and C. Faloutsos, Competing memes propagation on networks: A network science perspective, IEEE J. Sel. Areas Commun. **31**, 1049 (2013).

[9] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen, The spread of awareness and its impact on epidemic outbreaks, Proc. Natl. Acad. Sci. USA **106**, 6872 (2009).

[10] V. Marceau, P. A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, Modeling the dynamical interaction between epidemics on overlay networks, Phys. Rev. E **84**, 026105 (2011).

[11] P. Jaccard, Étude comparative de la distribution florale dans une portion des alpes et des Jura, Bull. Soc. Vaudoise Sci. Nat. **37**, 547 (1901).

[12] V. Havel, A remark on the existence of finite graphs, Časopis Pěst. Mat. (in Czech) **80**, 477 (1955).

[13] S. L. Hakimi, On realizability of a set of integers as degrees of the vertices of a linear graph. I, J. Soc. Ind. Appl. Math. **10**, 496 (1962).

[14] T. Britton, M. Deijfen, and A. Martin-Löf, Generating simple random graphs with prescribed degree distribution, J. Stat. Phys. **124**, 1377 (2006).

[15] E. A. Bender and E. R. Canfield, The asymptotic number of labeled graphs with given degree sequences, J. Comb. Theory, Ser. A **24**, 296 (1978).

[16] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labeled regular graphs, Eur. J. Combin. **1**, 311 (1980).

[17] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, Random Struct. Algorithms **6**, 161 (1995).

[18] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E **64**, 026118 (2001).

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

[20] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Generation of uncorrelated random scale-free networks, Phys. Rev. E **71**, 027103 (2005).

[21] R. Xulvi-Brunet and I. M. Sokolov, Changing correlations in networks: Assortativity and dissortativity, Acta Phys. Pol. B **36**, 1431 (2005).

[22] M. E. J. Newman, Assortative Mixing in Networks, Phys. Rev. Lett. **89**, 208701 (2002).

[23] M. G. Kendall, A new measure of rank correlation, Biometrika **30**, 81 (1938).

[24] G. Bell and J. Potterat, Partner notification for sexually transmitted infections in the modern world: A practitioner perspective on challenges and opportunities, Sex. Transm. Infect. **87**, ii34 (2011).

[25] D. Juher, J. Saldaña, R. Kohn, K. Bernstein, and C. Scoglio, Network-centric interventions to contain the syphilis epidemic in san francisco, Sci. Rep. **7**, 6464 (2017).

[26] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Wiley Online Library, 1992), Vol. 28.

[27] R. M. May and A. L. Lloyd, Infection dynamics on scale-free networks, Phys. Rev. E **64**, 066112 (2001).

[28] R. Pastor-Satorras and A. Vespignani, Epidemic Spreading in Scale-Free Networks, Phys. Rev. Lett. **86**, 3200 (2001).

[29] J. C. Miller, A. C. Slim, and E. M. Volz, Edge-based compartmental modeling for infectious disease spread, J. R. Soc. Interface **9**, 890 (2012).