# Epidemic Spreading in Group-Structured Populations

Siddharth Patwardhan[1,*] Varun K. Rao[1,2,*] Santo Fortunato[1] and Filippo Radicchi[1]

[1]*Center for Complex Networks and Systems Research, Luddy School of Informatics,*
*Computing, and Engineering, Indiana University, Bloomington, Indiana 47408, USA*
[2]*School of Public Health, Indiana University, Bloomington, Indiana 47408, USA*

Individuals involved in common group activities or settings, e.g., college students that are enrolled in the same class and/or live in the same dorm, are exposed to recurrent contacts of physical proximity. These contacts are known to mediate the spread of an infectious disease; however, it is not obvious how the properties of the spreading process are determined by the structure of and the interrelation among the group settings that are at the root of those recurrent interactions. Here, we show that reshaping the organization of groups within a population can be used as an effective strategy to decrease the severity of an epidemic. Specifically, we show that when group structures are sufficiently correlated, e.g., the likelihood for two students living in the same dorm to attend the same class is sufficiently high, outbreaks are longer but milder than for uncorrelated group structures. Also, we show that the effectiveness of interventions for disease containment increases as the correlation among group structures increases. We demonstrate the practical relevance of our findings by taking advantage of data about housing and attendance of students at the Indiana University campus in Bloomington. By appropriately optimizing the assignment of students to dorms based on their enrollment, we are able to observe a twofold to fivefold reduction in the severity of simulated epidemic processes.

Subject Areas: Complex Systems,
Interdisciplinary Physics,
Statistical Physics

## I. INTRODUCTION

From the Black Plague in the 13th century to the recent COVID-19 pandemic, epidemics have always represented significant threats for humanity [1]. Network science has been at the core of many advances made in epidemic modeling, given that the fate of an epidemic process is fundamentally determined by the structure of the social network where spreading occurs [2–5]. Notably, network structural properties play nontrivial roles in shaping not only the dynamics of spreading, but also mitigation strategies [2,6–10].

Contacts of physical proximity that mediate spreading among individuals often occur in group activities or settings [5,11,12]. For example, in a college, two students can get in contact because they are enrolled in the same class and/or live in the same dorm. Similarly, physical proximity between school-going children happen in

classroom or family settings. Realistic models of epidemic spreading are informed by data accounting for the existence of multiple types of group interactions that individuals are exposed to [13]. This aspect is also accounted for in some recent theoretical metapopulation models of epidemic spreading [5,11,12,14–16].

A common feature of the above-mentioned studies is considering the exposure of individuals to multiple types of group interactions as an input rather than a free, or tunable, parameter of the spreading model. For example, Granell and Mucha assume that the groups of a population are given and then analytically study how the threshold value of an epidemic spreading in the group-structured population depends on the mobility of individuals among groups [11]. In this paper, we change perspective by focusing our attention on the effects that group structure and interrelation among groups have on the properties of a susceptible-infected-recovered epidemic process. To this end, we represent interaction patterns among individuals of a population as edge-colored graphs [17] with block or community structure [18]. Each color or layer in these graphs represents a specific social setting where two individuals may get in physical proximity, e.g., a classroom or a dorm in the case of college students. Blocks or communities are instead used to model group interactions

---

in the layers. In the example of a population of college students, blocks represent specific classes attended by students or specific dorms where students reside. The framework allows us to control for the strength of the block structure of the layers, e.g., the likelihood that two students in the same class or dorm have a proximity contact, as well as for the similarity of groups across layers, e.g., the likelihood that two students live in the same dorm and attend the same class. We study the importance of these two factors in determining the fate of epidemic processes.

A study that is immediately related to ours is the one by Fan et al. [19]. They consider epidemic spreading in edge-colored graphs that are embedded in geometric space. They numerically estimate the epidemic threshold for graphs with variable levels of correlation between the layers' embeddings, and find that such a quantity grows as geometric correlation increases. In other words, the severity of an epidemic is reduced in a network with geometrically correlated layers compared to a network whose layers are not correlated.

As there exists a tight analogy between geometric embedding and community structure [20], the work by Fan et al. directly relates to ours. In our work, however, we go well beyond the state of the art by expanding the analysis by Fan et al. in two main, fundamental respects. First, we provide a full characterization of the importance of group strength and correlation in epidemic spreading by monitoring not only static quantities such as the size of the outbreak and the epidemic threshold, but also dynamical quantities such as the duration and intensity of the epidemic. Looking at dynamical metrics of epidemic severity turns out to be essential for a full understanding of how topological correlations affect spreading dynamics. In particular, there are cases where changes in group correlation do not lead to any variations in outbreak size and epidemic threshold, but to apparent variations of both duration and intensity. In addition, we consider the role of correlation in interventions for disease containment, an aspect that is not considered by Fan et al., but that actually displays a marked dependence on the topological correlations that characterize a network. We find that herd immunity in a network with correlated group structure can be achieved by immunizing a fraction of individuals that is even fivefold smaller than in the case of a network with noncorrelated group structure. Second, and more important, Fan et al. do not apply their approach to the analysis of real-world systems. The method is in fact limited by a series of challenges in practical applications. The approach requires knowledge of the network topology of the layers to perform the embedding in the hyperbolic space, but detailed network data are in many cases not readily available. Also, once the embedding is performed, manipulating the network to reduce the severity of a disease outbreak is not an intuitive task as it requires changing

coordinates of nodes in the hyperbolic space. Our approach instead does not require knowledge of the network of contacts. We simply use information about groups' composition. This information is generally available in real systems. Further, interventions on the network are easy to formulate or implement as they require only changing the group assignments of nodes. In this paper, we actually demonstrate the utility of our proposed framework in the analysis of a real-world scenario of data concerning enrollment and housing of college students at the Indiana University campus in Bloomington. We show that a simple strategy of reassigning students to housing facilities can (i) dramatically reduce the severity of an epidemic outbreak, both in terms of intensity and duration, as well as (ii) greatly enhance the effectiveness of interventions for disease containment.

## II. RESULTS

### A. Synthetic graphs

In Fig. 1, we provide an example of our modeling framework, see Sec. IV for details. Continuous-time susceptible-infected-recovered (SIR) dynamics happens on an edge-colored graph with $N$ nodes [2,17]. For simplicity, in most of our analysis we assume that there are only two layers of interactions. To have a concrete example in mind, think of a population of college students, where layer $\ell = 1$ represents interactions in housing facilities and layer $\ell = 2$ represents classroom interactions. Individual layers of the network have preimposed block or community structure constructed according to the rules of the stochastic block model (SBM) [18]. Once more for simplicity, we assume the model to be homogeneous both in terms of degree and block sizes. Specifically, nodes' degrees in layer $\ell$ obey a Poisson distribution with average $\langle k^{(\ell)} \rangle$. Also, each individual $i$ is associated to a group $\sigma_i^{(\ell)} = 1, \ldots, Q^{(\ell)}$ in layer $\ell$, with all groups having the same size $q^{(\ell)}$. The mixing parameter $\mu$ quantifies the fraction of connections that each node has toward other nodes outside its own group. Please note that we use the same mixing parameter for both layers just for convenience, but this is not a requirement of our model. The correlation between the community structure of two layers is measured in terms of normalized mutual information (NMI) [21]. Perfectly aligned communities generate large NMI values, whereas uncorrelated block structures have low NMI values. Concerning SIR dynamics, the recovery rate is set equal to one, whereas spreading events occur in both layers at rate $\beta$. The choice of using identical spreading rates for both layers is made for simplicity; the effective rate of spreading is anyway also a function of $\langle k^{(\ell)} \rangle$. Initial conditions of the dynamics are such that all nodes are in the susceptible state, except for a single random node that is set in the infected state.
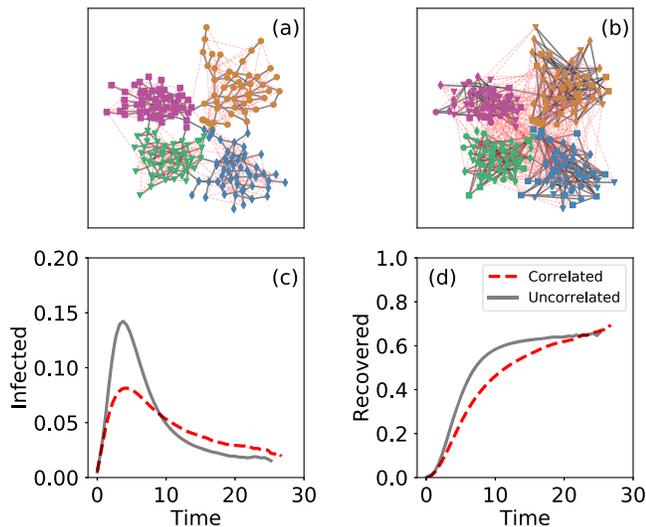
FIG. 1. Modeling framework and metrics of epidemic severity. (a) As a basic illustration of the modeling framework, we consider an edge-colored graph composed of $N = 200$ nodes and two layers of interactions. Nodes are organized in four groups of size $q^{(\ell)} = 50$ for both layer $\ell = 1$ and $\ell = 2$. Colors of the nodes (i.e., orange, purple, blue, green) represent the community structure in layer $\ell = 1$; shapes (i.e., circle, square, triangle, diamond) indicate the community memberships in layer $\ell = 2$. Here, the community structures of the layers are maximally correlated (NMI = 1.0, with NMI standing for normalized mutual information), meaning that there is a one-to-one map from shapes to colors. Connections among pairs of nodes are created such that the average degree is $\langle k^{(\ell)} \rangle = 3$; a fraction $\mu = 0.025$ of these edges connects a node to other nodes outside its own community. Using these parameters, we generate edges in layer $\ell = 1$ (full gray lines) and in layer $\ell = 2$ (dashed red lines). (b) Same as in (a), but for uncorrelated community structure (NMI = 0.0); i.e., colors and shapes are assigned to nodes randomly. (c) We run $V = 1\,000\,000$ simulations of the SIR model on top of the graphs of panels (a) and (b) by setting the spreading rate $\beta = 0.4$, i.e., setting the basic reproduction number $R_0 = \beta(\langle k^{(1)} \rangle + \langle k^{(2)} \rangle) = 2.4$. We display the fraction of infected nodes in the population as a function of time. We consider bins of size 0.05 and report only the average values over at least 100 surviving runs. (d) Same as in (c), but here we are displaying the fraction of recovered nodes in the population as a function of time.

In the example of Fig. 1, we play with only one ingredient, that is the correlation of the community structure of the layers. Individual layers have a neat block structure, as they are generated for $\mu = 0.025$. However, the two structures are in one case maximally correlated (NMI = 1.0) and in the other case completely uncorrelated (NMI = 0.0). In the example of the population of college students, correlated blocks occur when the conditional probability for two students to attend the same class given that they live in the same dorm is higher than the probability of two random students to attend the same class;

uncorrelated community structure indicates that the two probabilities are equal.

The effect that correlation has on the outcome of SIR spreading is particularly interesting even for a small population like the one considered in Fig. 1. Outbreaks are comparable in size between the correlated and uncorrelated cases; however, graphs with correlated community structure display milder and longer epidemics than graphs with uncorrelated groups. In essence, topological correlations among the layerwise community structure of the network may have an impact on how the epidemic unfolds, but not on its magnitude. This is a fundamentally important aspect that was not captured in the analysis by Fan *et al.* [19].

We study in a systematic manner the effect that the correlation among the community structure of the graph layers has on SIR spreading. To this end, we employ a simple algorithm that allows us to generate block structures anywhere between the maximally correlated and the uncorrelated configurations, see Sec. IV for details. In Fig. 2, we display results for a graph with $N = 10\,000$ nodes. Model parameters are $\langle k^{(1)} \rangle = 3$, $\langle k^{(2)} \rangle = 10$, $q^{(1)} = 5$, $q^{(2)} = 25$, and $\mu = 0.025$. We consider three distinct values of the SIR spreading rate, $\beta = 0.2$, 0.4 and 0.6, all of them corresponding to the endemic regime of the dynamics. As the figure clearly displays, increasing the correlation between the block structure of the layers reduces the severity of the epidemic. There is no apparent reduction in terms of outbreak size, unless the layer community partitions are very close to the configuration corresponding to maximum correlation. Severity reduction is instead mostly visible in terms of intensity and duration. Essentially, spreading slows down as the correlation of the partitions is increased, with a clear reduction in the peak value of the fraction of infected and an apparent increase in the total duration of the epidemic.

The above findings are valid for network layers that are sufficiently modular. When the strength of the community structure varies, nothing happens if the layers' partitions are uncorrelated, see Figs. 2(d)–2(f). However, if partitions are correlated, we observe an increase of epidemic severity as the community structure progressively becomes loose. Further, we observe an interesting trend in the metrics of epidemic severity as the number or size of communities is varied [see Figs. S1 and S2 in Supplemental Material (SM) [22] ]. The trend is visible only if partitions are correlated, and community structure is sufficiently strong. In such a case, the outbreak size, the duration, and the peak of infected nodes grow as the size of the blocks decreases. Once more, if partitions are not correlated, then no change in the values of metrics of epidemic severity is visible as the size or number of clusters is varied. All the above considerations are still valid even if we consider models with heterogeneous community sizes (Fig. S3 [22]) and node degrees (Fig. S4 [22]). Further, results for edge-colored graphs composed of three layers are almost
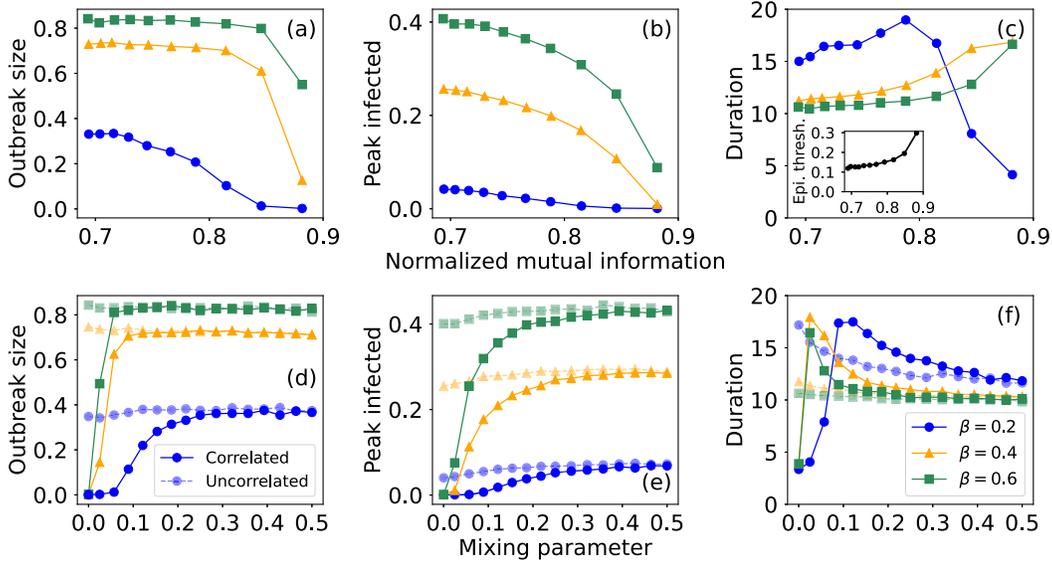
FIG. 2. Epidemic spreading in synthetic group-structured populations. (a) We consider edge-colored graphs with $N = 10\,000$ nodes and parameters $\langle k^{(1)} \rangle = 3$, $q^{(1)} = 5$, $\langle k^{(2)} \rangle = 10$, $q^{(2)} = 25$, and $\mu = 0.025$. We tune the correlation among the community structure of the layers by swapping community memberships of nodes as explained in Sec. IV. We simulate SIR dynamics, and measure the size of the outbreak. We plot it as a function of the NMI between the layers' partitions. Results are averaged over $V = 5000$ repetitions. Different colors and symbols refer to results valid for different choices of the spreading rate $\beta$, i.e., $\beta = 0.2$, 0.4, and 0.6, all corresponding to supercritical spreading rates [see inset in (c)]. The values of the reproduction number are $R_0 = 2.6$, 5.2, and 7.8, respectively. (b) Same as in (a), but for the peak value of the fraction of infected. (c) Same as in (a), but for the duration of the epidemic. (d) We consider only the configurations corresponding to maximum (full curves, solid symbols) and minimum (dashed curves, transparent symbols) correlation among layers' partitions, and generate networks with variable mixing parameter $\mu$. We plot the size of the outbreak as function of $\mu$. Results represent averages over $V = 5000$ realizations of the model. (e) Same as in (d), but for the peak of the fraction of infected. (f) Same as in (d), but for the average duration of the epidemic.

identical to those obtained on two-layer edge-colored graphs (Fig. S5 [22]).

Also, we analyze how the two main ingredients of our network model, i.e., the strength of the community structure of the individual layers and the correlation among the community partitions of the layers, influence the effectiveness of immunization in suppressing an epidemic, see Fig. 3. To this end, we simply change the initial conditions of the dynamics by imposing that a fraction of randomly chosen nodes is set to the recovered state. All other nodes are in the susceptible state, except for one randomly chosen initial spreader that is set in the infected state. The effectiveness of random immunization is greatly enhanced when communities are correlated: at parity of spreading rate $\beta$, correlating the group structures leads to even a fivefold reduction in outbreak size and peak of
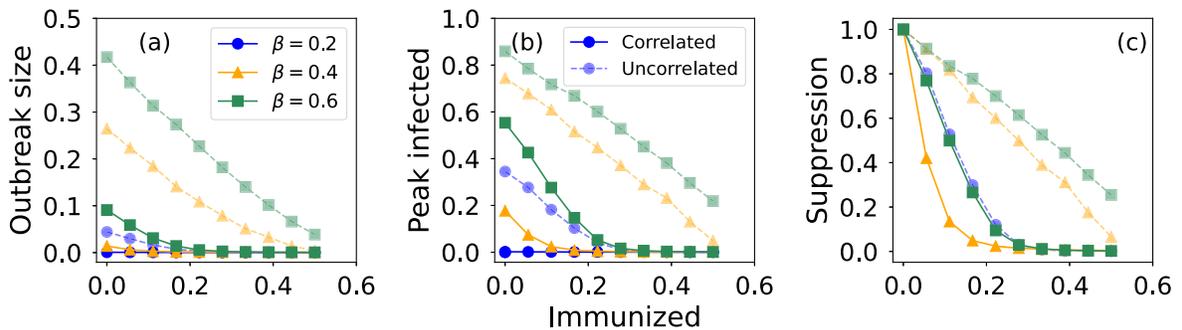


FIG. 3. Immunization in synthetic group-structured populations. (a) We consider the same experimental setting as in the bottom row of Fig. 2. We set the mixing parameter $\mu = 0.025$ and consider three values of the spreading rate $\beta$, i.e., 0.2, 0.4, and 0.6, corresponding to the reproduction number $R_0 = 2.6$, 5.2, and 7.8, respectively. We change, however, the initial condition of the dynamics, by immunizing a random fraction of nodes. We then plot the size of the outbreak as a function of the fraction of immunized nodes. Results are obtained by averaging the outcome of $V = 5000$ repetitions of the epidemic process. (b) Same as in (a), but for the peak fraction of infected nodes. (c) We rescale the abscissa values of (a) by the outbreak size that is observed when a null fraction of nodes is immunized.

infection compared to the uncorrelated case; also, the overall behavior of the system obtained for $\beta = 0.6$ in the presence of group correlation is almost identical to the one observable for $\beta = 0.2$ in a population with uncorrelated group structure. A similar conclusion is reached by measuring the site-percolation thresholds of the networks, see Fig. S6 [22]. The site-percolation model is closely related to immunization, as the immunized nodes and their connections are effectively removed from the system and the spreading dynamics is governed by the network formed by the edges between the nonimmunized nodes.

All the above results are obtained by means of numerical simulations. In the Supplemental Material [22], we describe the derivation of two theoretical approximations: the individual-based mean-field approximation (IBMFA) and the group-based mean-field approximation (GBMFA). IBMFA is a standard theoretical approach that uses as input the topology of the network [2]. GBMFA is an approximation introduced in this paper, inspired by the so-called degree-based mean-field approximation where nodes with identical degrees are treated as indistinguishable elements [2]. The main difference in GBMFA is that classes of indistinguishable nodes are defined based on the layer-wise community structures. IBMFA and GBMFA generate similar predictions; one of these predictions is that the

epidemic threshold equals $\beta_c = (\langle k^{(1)} \rangle + \langle k^{(2)} \rangle)^{-1}$ for any level of correlation existing between the layers' community structures, see Fig. S7 [22]. Qualitatively, this theoretical prediction supports our numerical findings on the weak dependence of the long-term metrics of epidemic severity from topological correlations among the network layers, highlighting once more fundamental differences existing between our results and those by Fan *et al.* [19]. From the quantitative point of view, however, we find the mean-field approximations to be accurate enough only for populations with sufficiently large communities or sufficiently large average degree, see Figs. S8–S11 [22]. These regimes, however, are of little importance in realistic settings where groups are generally much smaller than the size of the entire population and the average degree of individual nodes is small.

## B. College housing and attendance network

The paper has thus far presented simulations on synthetic group-structured populations aimed at understanding the fundamental principles of epidemic spreading in these systems. In this section, we demonstrate the significance and effectiveness of these principles in a real-world scenario. We examine data about housing and attendance
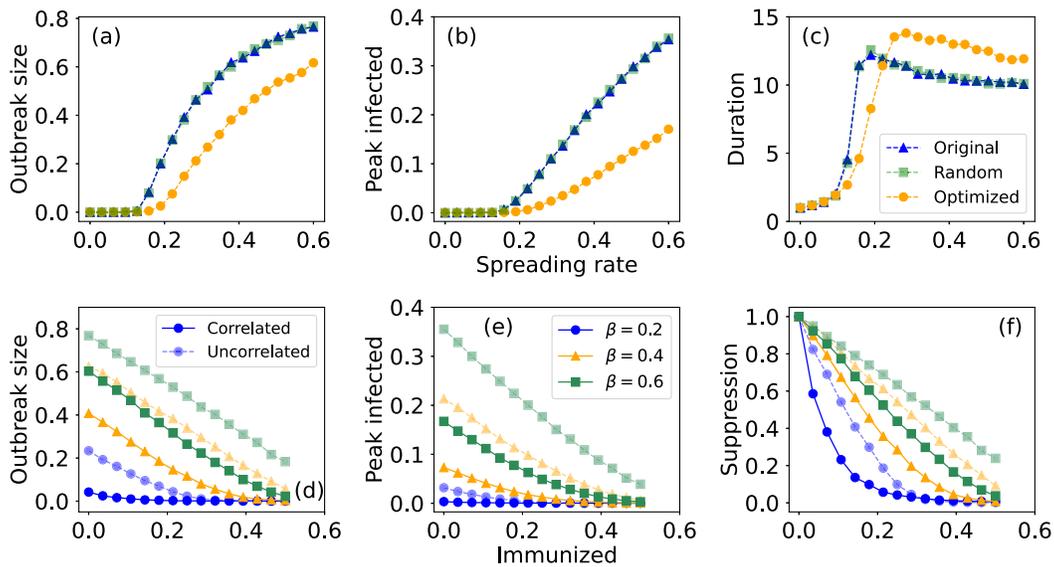


FIG. 4.  Epidemic spreading in the student population of the Indiana University Bloomington (IUB). (a) We use data about housing and attendance for the Fall 2019 semester at the IUB campus to generate edge-colored graphs with block structure. The community partition in one layer reflects housing assignments; the partition in the other layer serves to group students based on their program and education level. Different graphs are generated depending on whether network partitions are (i) those directly observed from the data, (ii) randomized, or (iii) optimized for maximum correlation. We then simulate SIR dynamics on the graphs and measure the average size of the outbreak as a function of the spreading rate $\beta$. Results are averaged over $V = 5000$ repetitions. (b) Same as in (a), but for the peak fraction of infected. (c) Same as in (a), but for the duration of the spreading process. (d) We plot the size of the outbreak as a function of the fraction of individuals that are initially immunized. We consider three values of the spreading rate $\beta$, i.e., 0.2, 0.4, and 0.6, corresponding to the reproduction number $R_0 = 2, 4$, and 6, respectively. Different symbols correspond to different $\beta$ values; full curves and solid symbols indicate the optimized configuration considered in (a); dashed curves and transparent symbols refer to graphs created using ground-truth partitions. Results are averaged over $V = 5000$ repetitions. (e) Same as in (d), but for the peak fraction of infected. (f) Same as in (d), but with abscissa values rescaled by the outbreak size observed when zero individuals are initially immunized.

at Indiana University Bloomington (IUB), see Sec. IV for details. There were $N = 10\,132$ individuals in the IUB dataset, each representing a student who resided in one of the campus facilities during the Fall 2019 semester. We use layer $\ell = 1$ to represent contacts between students in housing settings. We form $Q^{(1)} = 396$ groups composed of students living in the same floor of a large dormitory, or in the same Greek house. The average size of these groups is $\langle q^{(1)} \rangle = 25.58$. Layer $\ell = 2$ is used to represent classroom interactions. We form $Q^{(2)} = 600$ groups of students based on their enrollment program (e.g., computer science, finance, mathematics, physics) and their education level (from freshman to Ph.D.). The average size of these groups is $\langle q^{(2)} \rangle = 16.86$. Note that the partitions within the classroom and residential layers are disjoint. Moreover, the group sizes in the two partitions are not homogeneous, see Fig. S12 [22].

The ground-truth community partitions display a correlation level similar to that of the uncorrelated case. Specifically, the NMI between the partitions obtained from the data is 0.389, which is comparable with the one obtained by randomizing the dormitory assignment of students, which is 0.327. We also create an artificial configuration, where students are reassigned to residences so that the NMI between the partitions of the two layers is maximized. The optimization technique we employ is described in Sec. IV. The optimized configuration corresponds to the NMI value 0.934.

Given the three configurations described above, we generate modular network layers using the Lancichinetti-Fortunato-Radicchi (LFR) model [23], see Sec. IV for details. Nodes' degrees obey Poisson distributions with average equal to $\langle k^{(1)} \rangle = \langle k^{(2)} \rangle = 5$. The strength of the community structure is controlled by the value of the mixing parameter, set here to $\mu = 0.025$. We stress that the sizes of the communities in the two layers is always set equal to the one learned directly from the real data.

In Figs. 4(a)–4(c), we display the metrics of epidemic severity as functions of the spreading rate. No difference between the ground-truth and randomized configurations is noticeable. The optimized configuration displays instead a less severe epidemic compared to the other two configurations. Also, we notice that random immunization becomes more effective when communities are correlated than when no correlation among the layers' partitions is present, Figs. 4(d)–4(f). Qualitatively similar results are obtained also for $\mu = 0.1$, see Fig. S13 [22].

## III. DISCUSSION

In this paper, we studied SIR spreading in group-structured populations. The real system that inspired our work is a population of college students that reside on campus during an academic semester. Recurrent interactions among students are due to group activities or

settings, e.g., housing and class enrollment; each setting is responsible for the formation of network layers with modular structure. At the level of individual layers, modular structure is characterized by the strength and size of the blocks or communities that are present in the layer. In the example of the population of college students, these correspond to the size of classes or dorms, and to the level of mixing between students depending on their enrollment or housing assignment. Correlation among the layerwise community structures is indicative instead for the likelihood of two nodes to belong to the same community in both layers, i.e., the conditional probability of two students to attend the same class given that they live in the same dorm. All these parameters play fundamental roles for the determination of properties characterizing the spread of an infectious disease within the population.

In the first part of the paper, we considered synthetic populations where the above ingredients can be finely tuned. By means of extensive numerical simulations, we showed that SIR dynamics is characterized by mild and long spreading processes if the communities of the individual layers are neat, and the correlation among the layerwise community structure is sufficiently high. On the contrary, weak and/or uncorrelated community structures generally correspond to short and intense outbreaks. Immunization in populations with correlated communities is also greatly more effective than in populations with uncorrelated layerwise community structures.

In the second part of the paper, we took advantage of real data about housing and attendance of students at the Indiana University Bloomington campus, and numerically simulated SIR dynamics in this population. Ground-truth data correspond to layerwise partitions that are uncorrelated. By appropriately reassigning students to housing facilities on the basis of their enrollment profiles, we were able to show that the severity of an epidemic can be significantly reduced compared to the case in which no reassignment is performed. The reassignment strategy also enhances the effectiveness of random immunization in suppressing disease spreading.

The above results are obtained under the assumption that the layerwise group structure is sufficiently neat, i.e., for sufficiently low values of the mixing parameter $\mu$. This assumption can be empirically justified by analyzing publicly available datasets concerning contact networks within group-structured populations. First, we consider a contact network for a primary school. The dataset consists of events of physical proximity between students and teachers [24]. Each event of physical proximity is associated to a duration. The system contains 242 individuals and 11 communities (10 classrooms plus teachers). Two days of observation are included in the dataset: day one consists of 37414 contact events and day two of 40188 contact events. We find that the values of the weighted mixing parameter $\langle \mu \rangle$, i.e., the average fraction of contact

time spent by the students outside their classrooms, are $\langle \mu \rangle = 2.3 \times 10^{-5}$ and $\langle \mu \rangle = 2.6 \times 10^{-5}$, respectively. Next, we look at a dataset concerning physical proximity in an office building [25]. The dataset consists of 9827 contact events between 92 individuals organized in 5 departments (communities). We find $\langle \mu \rangle = 9.1 \times 10^{-4}$.

Our findings stem from the basic principle that the severity of an epidemic is proportional to the overall level of mixing between individuals of a population. Correlated modular structure in our framework corresponds to redundant interactions among subsets of individuals, which speeds up the propagation of a disease within a subset, but also slows down the diffusion across subsets thus in the overall population. Although having a radically different implementation, the mechanism of reducing the severity of an epidemic by enhancing group correlation is similar to the one at the basis of the mitigation strategy proposed by Meidan *et al.*, consisting of an alternating lockdown policy that involves half of the population at a time [26].

A very important remark is that our findings are apparent only from numerical simulations of SIR dynamics, but they do not emerge from the numerical integration of mean-field equations aimed at approximating SIR dynamics. The mismatch between ground-truth dynamics and its approximations is particularly apparent in realistic settings where groups are small compared to the population size; also, the mismatch mostly concerns early and mid stages of SIR dynamics, rather than its long-term features.

All our results are based on important simplifications concerning the spreading dynamics of realistic diseases as well as the contact networks that are the basis of the spread of infectious diseases within real-world populations. Nonetheless, we believe they provide some easy-to-implement principles to reduce the severity of real epidemics that managers can take under consideration when planning group activities in schools, colleges, or other large organizations.

## IV. METHODS

### A. Edge-colored block-structured graphs

We consider edge-colored graphs where $N$ nodes are connected via $L$ different types or layers of interactions [17]. We use $L = 2$ in most of our analysis; some results for $L = 3$ are presented in the SM [22]. The topology of the layer $\ell = 1, 2, \dots, L$ is fully specified by the adjacency matrix $A^{(\ell)}$. The generic element of this matrix $A_{ij}^{(\ell)} = A_{ji}^{(\ell)} = 1$ if nodes $i$ and $j$ are connected, and $A_{ij}^{(\ell)} = A_{ji}^{(\ell)} = 0$ otherwise.

#### 1. Synthetic graphs

In the majority of our experiments on synthetic graphs, network layers are instances of the version of the stochastic block model (SBM) [18] named planted partition model

[27]. Inputs for the generation of the network layer $\ell$ are the average degree $\langle k^{(\ell)} \rangle$, the mixing parameter $\mu^{(\ell)}$, the size $q^{(\ell)}$ of each community, and the vector $\vec{\sigma}^{(\ell)} = (\sigma_1^{(\ell)}, \sigma_2^{(\ell)}, \dots, \sigma_i^{(\ell)}, \dots, \sigma_N^{(\ell)})$ denoting the community membership of individual nodes, i.e., node $i$ belongs to community $\sigma_i^{(\ell)} = 1, \dots, Q^{(\ell)}$, with $Q^{(\ell)}$ total number of blocks in layer $\ell$; the output is the adjacency matrix $A^{(\ell)}$. The element $A_{ij}^{(\ell)}$ is a Bernoulli random variate equal to one with probability $p_{in}^{(\ell)} = \langle k^{(\ell)} \rangle (1 - \mu^{(\ell)})/(q^{(\ell)} - 1)$ if $\sigma_i^{(\ell)} = \sigma_j^{(\ell)}$, and with probability $p_{out}^{(\ell)} = \langle k^{(\ell)} \rangle \mu^{(\ell)}/(N - q^{(\ell)})$ if $\sigma_i^{(\ell)} \neq \sigma_j^{(\ell)}$. Note that this is the only model used to generate edge-colored graphs with $L = 3$ layers.

In some of our experiments on synthetic graphs, we generate network layers using the Lancichinetti-Fortunato-Radicchi (LFR) model [23], i.e., a variant of the SBM designed to deal with heterogeneous community sizes and/or nodes' degrees. More specifically, in the set of experiments where we consider power-law distributed community sizes, we first generate the community sizes $q_1^{(\ell)}, q_2^{(\ell)}, \dots, q_{Q^{(\ell)}}^{(\ell)}$. These are integer numbers randomly extracted from the distribution $P(q) \sim q^{-\tau}$ for $q \in [q_{min}, q_{max}]$ and $P(q) = 0$ otherwise, with $\tau$, $q_{min}$, and $q_{max}$ input parameters of the model. Clearly, we have that $\sum_{i=1}^{Q^{(\ell)}} q_i^{(\ell)} = N$. Please note that the same sequence for the community sizes is used for all layers of the network. Nodes are assigned to communities according to their sizes to form the community vector $\vec{\sigma}^{(\ell)}$. For the generation of the topology of the graph layer $\ell$, we first set the degree of each node to be a random variate extracted from a Poisson distribution with average $\langle k^{(\ell)} \rangle$. We then connect pairs of nodes at random, taking care of preserving for each node its preassigned degree value, as well as the ratio $\mu^{(\ell)}$ between intercommunity and total connections. Here the average degree $\langle k^{(\ell)} \rangle$ and the mixing parameter $\mu^{(\ell)}$ are input parameters of the model. In the set of experiments where we consider power-law degree distributions, the size of all communities in layer $\ell$ is equal to $q^{(\ell)}$, an input parameter of the model. We then assign degrees to nodes by extracting random numbers from the power-law distribution $P(k) \sim k^{-\gamma}$ for $k \in [k_{min}, k_{max}]$ and $P(k) = 0$ otherwise, with $\gamma$, $k_{min}$, and $k_{max}$ input parameters of the model. The degree sequence of each layer is generated independently. We then connect pairs of nodes at random, taking care of preserving for each node its preassigned degree value, as well as the ratio $\mu^{(\ell)}$ between intercommunity and total connections, with the mixing parameter $\mu^{(\ell)}$ being an input parameter of the model.

#### 2. College housing and attendance network

Also, we consider a dataset containing information about class attendance of students who resided in the Indiana

University Bloomington (IUB) campus during the Fall 2019 semester [28]. The dataset is composed of $N = 10\,132$ students. We use layer $\ell = 1$ to represent housing, and group together students living in the same floor of a dormitory building or in the same Greek house. We form $Q^{(1)} = 396$ groups of average size $\langle q^{(1)} \rangle = 25.58$. Please note that the size of the various groups is not constant, see Fig. S12 [22]. We use layer $\ell = 2$ to represent class interactions. Each student in the dataset is associated to one of the 218 degree programs (e.g., law, mathematics, physics) and 6 education levels (i.e., freshman, sophomore, junior, senior, master, and Ph.D.). On the basis of our data, we form $Q^{(2)} = 600$ blocks of average size $\langle q^{(2)} \rangle = 16.86$, by grouping together students belonging to the same degree program and education level. Note that not all possible compartments defined by the pairs degree program and education level contain at least one student. Also for layer $\ell = 2$, the size of the various groups is not a constant, see Fig. S12 [22]. Network layers considered in the analysis of the IUB dataset are generated according to the LFR model. Inputs required for the generation of the topology of layer $\ell$ are the average degree $\langle k^{(\ell)} \rangle$, the mixing parameter $\mu^{(\ell)}$, and the community-membership vector $\vec{\sigma}^{(\ell)}$. For the generation of the topology of the graph layer $\ell$, we first set the degree of each node to be a random variate extracted from a Poisson distribution with average $\langle k^{(\ell)} \rangle$. We then connect pairs of nodes at random, taking care of preserving for each node its preassigned degree value, as well as the ratio $\mu^{(\ell)}$ between intercommunity and total connections.

## B. Measuring and tuning correlation among the community structures of the layers

We measure the similarity between the community structure of two network layers, namely $\vec{\sigma}^{(1)}$ and $\vec{\sigma}^{(2)}$, in terms of normalized mutual information (NMI) [21]. By definition, NMI values are in the range [0, 1]; however, the actual minimum and maximum values that NMI can assume depend on the size and number of clusters in the two partitions. Small values of the NMI are associated to uncorrelated partitions, whereas large NMI values indicate correlated partitions.

### 1. Synthetic graphs

In our systematic tests on synthetic network models that are constructed using the SBM, we generate perfectly correlated partitions as follows. First, we fix the size $q^{(1)}$ of the clusters in layer $\ell = 1$, then we determine the size of the clusters in layer $\ell = 2$ as $q^{(2)} = dq^{(1)}$, where $d$ is a tunable integer parameter. For simplicity, we appropriately choose $q^{(1)}$ and $d$ so that both $q^{(1)}$ and $q^{(2)}$ are divisors of $N$. For example, in many of our tests we have $N = 10\,000$, $q^{(1)} = 5$, $d = 5$, and $q^{(2)} = 25$. Finally, we

set $\sigma_i^{(\ell)} = \lfloor i/q^{(\ell)} \rfloor + 1$ for all $i = 1, \ldots, N$ and for $\ell = 1, 2$, with $\lfloor \cdot \rfloor$ floor function.

In the tests where we take advantage of the LFR model, the size of the communities is identical in both layers. We thus generate perfectly correlated partitions by setting $\sigma_i^{(1)} = \sigma_i^{(2)}$ for all nodes $i$ in the network.

Starting from the partitions $\vec{\sigma}^{(1)}$ and $\vec{\sigma}^{(2)}$ that are maximally correlated, their NMI is reduced by considering each node, and swapping with probability $r$ its community membership in layer $\ell = 1$ with that of another randomly selected node. For $r = 0$, no swaps are actually performed so that the correlation of the community partitions is preserved; for $r = 1$, community labels in layer $\ell = 2$ are completely randomized compared to the initial configuration so that correlation between the layers' community memberships becomes minimal.

In the set of experiments where we consider edge-colored graphs composed of $L = 3$ SBM-generated network layers, all communities have the same size; thus we set $\sigma_i^{(1)} = \sigma_i^{(2)} = \sigma_i^{(3)}$ for all nodes $i$ in the network to generate perfectly correlated partitions. To decrease correlation, we consider each node and swap with probability $r$ its community membership in layers $\ell = 1$ and $\ell = 2$ with that of another randomly selected node. Note that we only consider the cases $r = 0$ and $r = 1$ for this specific set of experiments.

### 2. College housing and attendance network

To create completely uncorrelated partitions, we simply swap the housing assignment of each student with that of another randomly selected student.

To create a configuration with highly correlated partitions, we take advantage of the following greedy optimization technique. We define the tuple $\mathcal{P}^{(\ell)} = (z_1^{(\ell)}, \ldots, z_{Q^{(\ell)}}^{(\ell)})$, where $z_v^{(\ell)}$ denotes the remaining availability of group $v$ in layer $\ell$. We initially set $z_v^{(\ell)} = q_v^{(\ell)}$ for $v = 1, \ldots, Q^{(\ell)}$, where $q_v^{(\ell)}$, i.e., the size of group $v$ in layer $\ell$, is imposed by the IUB housing or attendance data. We set $M = 0$ denoting the number of elements that have been assigned to groups in the layers. We then iterate the following operations:

(1) We find the group label $r^{(\ell)}$ corresponding to the largest element in $\mathcal{P}(\ell)$. We can have either $r^{(1)} \geq r^{(2)}$ or $r^{(2)} \geq r^{(1)}$. We indicate with $r^{(\ell_t)}$ the larger of the two values, where $\ell_t$ is the label of the corresponding layer; we use $r^{(\ell_s)}$ to indicate the smaller of the two values, with $\ell_s$ denoting the label of the corresponding layer.

(2) We set $\sigma_i^{(\ell_s)} = r^{(\ell_s)}$ and $\sigma_i^{(\ell_t)} = r^{(\ell_t)}$ for all $i = M + 1, \ldots, M + z_{r^{(\ell_s)}}^{(\ell_s)}$; i.e., we assign $z_{r^{(\ell_s)}}^{(\ell_s)}$ nodes to coherent groups in both layers.

(3) We update $M \to M + z^{(\ell_s)}_{r^{(\ell_s)}}$, $z^{(\ell_t)}_{r^{(\ell_t)}} \to z^{(\ell_t)}_{r^{(\ell_t)}} - z^{(\ell_s)}_{r^{(\ell_s)}}$, and $z^{(\ell_s)}_{r^{(\ell_s)}} \to 0$.

(4) Until all the individuals have been assigned to a group, i.e., $M < N$, we go back to point 1, otherwise, we end the iterative procedure.

In the above step 2, we maximize the overlap of the largest groups still available in the layers. This greedy choice is iterated multiple times until all individuals have been assigned to groups in both layers. At the end of the algorithm, all groups of a given layer will have the desired size. Since the labels of the nodes are completely arbitrary, the obtained partitions can be thought as equivalent to reassigning students to dorms based on their enrollment.

## C. SIR dynamics on edge-colored graphs

We consider continuous-time SIR dynamics on an edge-colored graph [19]. Without loss of generality, we assume that the rate of recovery equals one. This is a standard setting as the behavior of SIR models is determined by the ratio of the spreading and recovery rates rather than their raw values. We indicate with $\rho_i(t) = S$, $I$, or $R$ the state of node $i$ at time $t$. Our initial conditions are generally such that all nodes are in the $S$ state, except for a randomly selected node $v$ that is in state $I$, i.e., $\rho_i(t = 0) = S$ for $i = 1, ..., N$ and $i \neq v$, while $\rho_v(t = 0) = I$. In some of our experiments, we also impose that a fraction $f$ of randomly selected nodes is in the recovered state, i.e., $\rho_i(t = 0) = R$ with probability $f$ for all $i = 1, ..., N$.

Starting from an initial condition, the rules of SIR dynamics are as follows. A generic node $i$ can change its state in the infinitesimal time interval $[t, t + dt]$ if (i) node $i$ is in the infected state and recovers with probability $dt$, i.e., $\rho_i(t) = I \to \rho_i(t + dt) = R$ with probability $dt$, (ii) node $i$ is in the susceptible state and gets infected with probability $\beta^{(\ell)} dt$ by its infected neighbor $j$ on layer $\ell$, i.e., $\rho_i(t) = S \to \rho_i(t + dt) = I$ with probability $\beta^{(\ell)} dt$ if $\rho_j(t) = I$ and $A^{(\ell)}_{ij} = 1$. Note that spreading events of type of (ii) can happen for all individual edges in all layers in the graph; thus if node $i$ is in the $S$ state at time $t$, the overall probability to get infected in the time interval $[t, t + dt]$ is $\mathrm{Prob}[\rho_i(t) = S \to \rho_i(t + dt) = I] = 1 - \prod_{\ell=1}^{L}(1 - \beta^{(\ell)} dt)^{g_i^{(\ell)}}$, where $g_i^{(\ell)} = \sum_j A^{(\ell)}_{ij} \delta(\rho_j(t), I)$ is the number of infected neighbors that node $i$ has in layer $\ell$ at time $t$ with $\delta(x, y) = 1$ if $x = y$ and $\delta(x, y) = 0$ otherwise. The dynamics proceeds until no nodes are in the infected state. The model is efficiently implemented via the Gillespie's algorithm [29].

### 1. Metrics of epidemic severity

We characterize the behavior of the SIR model on a given edge-colored graph by measuring the values of the following metrics:

(i) The outbreak size, i.e., the fraction of nodes that are in state $R$ at the end of the epidemic.

(ii) The peak of infection, i.e., the maximum value of the fraction of infected nodes that are simultaneously present in the network.

(iii) The epidemic duration, i.e., the instant of time when the last infected node recovers.

Given an edge-colored graph and the parameters $\beta^{(\ell)}$ of the epidemic model, we simulate multiple times SIR dynamics. Values of the above metrics displayed in our figures correspond to their sample averages.

In the inset of Fig. 2(c), we approximate the pseudo-critical epidemic threshold of the given edge-colored graph as the value of the spreading rate $\beta$ corresponding to the maximum of the ratio between the sample standard deviation and the sample average of the outbreak size.

The code and data used in this study are available at Ref. [30].

[1] Lizzie Wade, *From Black Death to fatal flu, past pandemics show why people on the margins suffer most*, Science (2020), https://www.science.org/content/article/black-death-fatal-flu-past-pandemics-show-why-people-margins-suffer-most.

[2] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani, *Epidemic processes in complex networks*, Rev. Mod. Phys. **87**, 925 (2015).

[3] Alessandro Vespignani, Huaiyu Tian, Christopher Dye, James O. Lloyd-Smith, Rosalind M. Eggo, Munik Shrestha, Samuel V. Scarpino, Bernardo Gutierrez, Moritz U. G. Kraemer, Joseph Wu *et al.*, *Modelling COVID-19*, Nat. Rev. Phys. **2**, 279 (2020).

[4] Quan-Hui Liu, Marco Ajelli, Alberto Aleta, Stefano Merler, Yamir Moreno, and Alessandro Vespignani, *Measurability of the epidemic reproduction number in data-driven contact networks*, Proc. Natl. Acad. Sci. U.S.A. **115**, 12680 (2018).

[5] David Soriano-Paños, L. Lotero, Alex Arenas, and Jesús Gómez-Gardeñes, *Spreading processes in multiplex meta-populations containing different mobility networks*, Phys. Rev. X **8**, 031039 (2018).

[6] Naoki Masuda, *Immunization of networks with community structure*, New J. Phys. **11**, 123018 (2009).

[7] Romualdo Pastor-Satorras and Alessandro Vespignani, *Immunization of complex networks*, Phys. Rev. E **65**, 036104 (2002).

[8] Yiping Chen, Gerald Paul, Shlomo Havlin, Fredrik Liljeros, and H. Eugene Stanley, *Finding a better immunization strategy*, Phys. Rev. Lett. **101**, 058701 (2008).

[9] Romualdo Pastor-Satorras and Alessandro Vespignani, *Epidemics and immunization in scale-free networks, Handbook of Graphs and Networks* (Wiley-VCH, Berlin, 2003).

[10] Saeed Osat, Fragkiskos Papadopoulos, Andreia Sofia Teixeira, and Filippo Radicchi, *Embedding-aided network dismantling*, Phys. Rev. Res. **5**, 013076 (2023).

[11] Clara Granell and Peter J. Mucha, *Epidemic spreading in localized environments with recurrent mobility patterns*, Phys. Rev. E **97**, 052302 (2018).

[12] Jesús Gómez-Gardeñes, David Soriano-Panos, and Alex Arenas, *Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks*, Nat. Phys. **14**, 391 (2018).

[13] Wouter Van den Broeck, Corrado Gioannini, Bruno Gonçalves, Marco Quaggiotto, Vittoria Colizza, and Alessandro Vespignani, *The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale*, BMC Infect. Dis. **11**, 1 (2011).

[14] Clara Granell, Sergio Gómez, and Alex Arenas, *Dynamical interplay between awareness and epidemic spreading in multiplex networks*, Phys. Rev. Lett. **111**, 128701 (2013).

[15] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas, *Diffusion dynamics on multiplex networks*, Phys. Rev. Lett. **110**, 028701 (2013).

[16] Clara Granell, Sergio Gómez, and Alex Arenas, *Competing spreading processes on multiplex networks: Awareness and epidemics*, Phys. Rev. E **90**, 012808 (2014).

[17] Frank P. Ramsey, *On a problem of formal logic, Classic Papers in Combinatorics* (1987), pp. 1–24, https://link.springer.com/chapter/10.1007/978-0-8176-4842-8_1.

[18] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt, *Stochastic blockmodels: First steps*, Soc. Networks **5**, 109 (1983).

[19] Dongmei Fan, Guo-Ping Jiang, Yu-Rong Song, and Xu Zhang, *Influence of geometric correlations on epidemic spreading in multiplex networks*, Physica A (Amsterdam) **533**, 122028 (2019).

[20] Ali Faqeeh, Saeed Osat, and Filippo Radicchi, *Characterizing the analogy between hyperbolic embedding and community structure of complex networks*, Phys. Rev. Lett. **121**, 098301 (2018).

[21] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas, *Comparing community structure identification*, J. Stat. Mech. (2005) P09008.

[22] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.13.041054 for additional numerical results and analytical derivations of the mean-field approximations.

[23] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi, *Benchmark graphs for testing community detection algorithms*, Phys. Rev. E **78**, 046110 (2008).

[24] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina *et al.*, *High-resolution measurements of face-to-face contact patterns in a primary school*, PLoS One **6**, e23176 (2011).

[25] Mathieu Génois, Christian L Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Barrat, *Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers*, Netw. Sci. **3**, 326 (2015).

[26] D. Meidan, N. Schulamann, Reuven Cohen, S. Haber, E. Yaniv, R. Sarid, and B. Barzel, *Alternating quarantine for sustainable mitigation of COVID-19*, Nat. Commun. **12**, 220 (2021).

[27] Anne Condon and Richard M Karp, *Algorithms for graph partitioning on the planted partition model*, in *Randomization, Approximation, and Combinatorial Optimization. Algorithms and Techniques: Third International Workshop on Randomization and Approximation Techniques in Computer Science, and Second International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (RANDOM-APPROX'99), Berkeley, CA, 1999* (Springer, New York, 1999), pp. 221–232.

[28] Gina Deom, Stefano Fiorini, Mark McConahay, Linda Shepard, and Julie Teague, *Data-driven decisions: Using network analysis to guide campus course offerings*, Coll. Univ. **96**, 2 (2021), https://eric.ed.gov/?id=EJ1324131.

[29] Daniel T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. **81**, 2340 (1977).

[30] S. Patwardhan, V. K. Rao, S. Fortunato, and F. Radicchi, *Code for epidemic spreading in group-structured populations* (2023), https://github.com/SiddharthP96/Multiplex-Epidemic-Spreading.