Journal of Complex Networks (2015) Page 1 of 44
doi:10.1093/comnet/cnv027

Null models for community detection in spatially embedded, temporal networks

MARTA SARZYNSKA[†]
Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
[†]Corresponding author. Email: sarzynska@maths.ox.ac.uk

ELIZABETH A. LEICHT
CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK

GERARDO CHOWELL
Mathematical, Computational & Modeling Sciences Center, School of Human Evolution and Social Change, Arizona State University, 900 S. Cady Mall, Tempe, AZ 85287-2402, USA

AND

MASON A. PORTER
Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK and CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK

Edited by: Danielle Bassett

[Received on 22 January 2015; accepted on 3 September 2015]

In the study of networks, it is often insightful to use algorithms to determine mesoscale features such as 'community structure', in which densely connected sets of nodes constitute 'communities' that have sparse connections to other communities. The most popular way of detecting communities algorithmically is to maximize the quality function known as modularity. When maximizing modularity, one compares the actual connections in a (static or time-dependent) network to the connections obtained from a random-graph ensemble that acts as a null model. The communities are then the sets of nodes that are connected to each other densely relative to what is expected from the null model. Clearly, the process of community detection depends fundamentally on the choice of the null model, so it is important to develop and analyse novel null models that take into account appropriate features of the system under study. In this paper, we investigate the effects of using null models that incorporate spatial information, and we propose a novel null model based on the radiation model of population spread. We also develop novel synthetic spatial benchmark networks in which the connections between entities are based on the distance or flux between nodes, and we compare the performance of static and time-dependent versions of the radiation null model to the standard ('Newman–Girvan') null model for modularity optimization and to a recently proposed gravity null model. In our comparisons, we use both the above synthetic benchmarks and time-dependent correlation networks that we construct using countrywide dengue fever incidence data for Peru. Our findings illustrate the need to use appropriate generative models for the development of spatial null models for community detection.

Keywords: community structure; modularity maximization; spatial null models; disease dynamics; radiation model; gravity model; synthetic benchmarks; time-dependent correlation networks.

1. Introduction

A network formalism is often very useful for describing complex systems of interacting entities [1,2]. Scholars in a diverse set of disciplines have studied networks for many decades, and network science has experienced particularly explosive growth during the last 20 years [1].

The most traditional network representation is an ordinary graph, in which nodes represent entities and edges represent pairwise connections between nodes. However, many networks are time-dependent [3,4] or multiplex (i.e., they include multiple types of connections between nodes) [5,6]. Moreover, network structure is influenced profoundly by spatial effects [7]. To avoid discarding potentially important information (as that can lead to misleading results), it is thus crucial to develop methods that incorporate features such as time-dependence, multiplexity and spatial embeddedness in a context-dependent manner [3,5,7]. Because of the newfound wealth of rich data, it has now become possible to validate increasingly complicated network structures and methods using empirical data.

In the present paper, we study a mesoscale network feature known as *community structure*. A 'community' is a set of nodes with dense connections among themselves, and with only sparse connections to other communities in a network [8,9]. Communities arise in numerous applications. For example, social networks typically include dense sets of nodes with common interests or other characteristics [10], networks of legislators often contain dense sets of individuals who vote in similar ways [11] and protein interaction networks include dense sets of nodes that are related to functional units [12]. The algorithmic detection of communities and the subsequent investigation of both their aggregate properties and the properties of their component nodes can provide novel insights into the relationship between network structure and function (e.g., functional groupings of newly discovered proteins [13]).

Myriad community-detection methods have been developed [8,9]. The most popular family of methods entails the optimization of a quality function known as *modularity* [14,15]. To maximize modularity, one compares the actual network structure to some *null model*, which quantifies what it means for a pair of nodes to be connected 'at random'. Traditionally, most studies have randomized only network structure (while preserving some structural properties) and not incorporated other features (such as spatial or other information). The standard null model for modularity optimization is the 'Newman–Girvan' (NG) null model, in which one randomizes edge weights such that the expected strength distribution is preserved [14,15]. It is thus related to the classical configuration model [1]. The NG null model is very popular because of its simplicity and effectiveness, and it has been derived systematically through the consideration of Laplacian dynamics on networks [16,17]. However, the NG null model is also a naive choice, as it does not incorporate domain-specific information.

The choice of a null model is an important consideration because (1) it can have a significant effect on the community structure obtained via optimization of a quality function and (2) it changes the interpretation of communities [18–20]. The best choice for a null model depends both on one's data set and on one's scientific question. In the present paper, we explore the issue of null-model choice in detail in the context of spatially embedded and temporal networks.

Most existing research on community detection does not incorporate metadata about nodes (or edges) or information about the timing and location of interactions between nodes. With the increasing wealth of space-resolved and time-resolved data sets, it is important to develop community-detection techniques that take advantage of the additional spatial and temporal information (and of domain-specific information, such as generative models for human interactions [21]). Indeed, community detection in temporal networks has become increasingly popular [22–28], but the majority of methods use networks that are constructed from either static snapshots of data or aggregations of data over time windows. Few investigations of community structure in temporal networks have used methods that exploit

temporal structure (see, e.g., [25,28]). There is also starting to be more work on the influence of space on community structure [21,29–32], but much more research is necessary.

In the present paper, we use modularity maximization to study communities in spatially embedded and time-dependent networks. We compare the results of community detection using two different spatial null models—a *gravity null model* [21] and a new *radiation null model*—to the standard NG null model using (1) novel synthetic benchmark networks that incorporate spatial effects via distance decay or disease flux and (2) temporal correlation networks that we construct using time-series data of recurrent epidemic outbreaks in Peru.

Our direct analysis of disease data in the present paper provides a complementary approach to the majority of studies in epidemiology that use network methodology. These other studies often focus on the importance of interpersonal contact networks on disease spread on an individual level, and their methodology and perspective have become increasingly prevalent in the modelling of infectious diseases [33]. Our work also complements other approaches, such as large-scale compartmental models that incorporate transportation networks to link local populations. Such compartmental models have been used to study large-scale spatial disease spread, and they have yielded insights into the influence of features such as spatial location, climate and facility of transportation on phenomena such as disease persistence and synchronization of disease spread [7,34–36].

The rest of our paper is organized as follows. In Section 2, we give an overview of networks and community detection. We also discuss the gravity null model and introduce a new radiation null model. We give our results for synthetic spatial networks in Section 3, and we give our results for correlation networks that we construct from disease data in Section 4. We summarize our results in Section 5. In the appendices, we include a comparison of the similarity measures that we use for spatial benchmarks with a ground-truth partition, and we perform additional numerical experiments for these benchmarks.

2. Networks and community structure

A network describes a set of entities (called *nodes*) that are connected by pairwise relationships (called *edges*). In the present paper, we study weighted networks that are *spatially embedded*; each node represents a location in space. One can represent a weighted network with N nodes as an $N \times N$ adjacency matrix W, where an edge W_{ij} represents the strength of the relationship between nodes i and j. We seek to find *communities*, which are sets of nodes that are densely connected to each other but sparsely connected to other dense sets in a network [8,9].

We wish to study the evolution of network structure through time. The simplest way to study temporal data is through an ordered set of *static networks*, which can arise as snapshots at different points in time or as a sequence of aggregations over consecutive time windows (which one can take either as overlapping or non-overlapping). See [3,20,37] for further discussion.

Static networks provide a good starting point for the development and investigation of new methods—which, in our case, entails how to incorporate spatial information into null models for community detection via modularity maximization. However, they do not take full advantage of temporal information in data that changes in time. For example, it can be hard to track the identity of communities in temporal sequences of networks [25].

To mitigate the problem of community tracking, we also use a type of *multilayer network* [5,6] known as a 'multislice network' [25]. Such a network can be represented using an $N \times N \times m$ adjacency tensor $\bar{W}$ that has m layers and N nodes in each layer, and we suppose that each layer has a copy of each node i (though this can be relaxed). The intralayer edges in the network are exactly the same as they are for the sequence of static networks: the tensor element $\bar{W}_{ijs}$ gives the weight of an intralayer

edge between nodes i and j in layer s. Additionally, each node j is adjacent to copies of itself in consecutive layers s and r using interlayer edges of weight C_{jsr}. In this paper, we assume for simplicity that $C_{jsr} = \omega \in [0, \infty)$ for all j, s and r, but one can also consider more general situations [5,38]. A multislice network can have up to Nm multilayer nodes (i.e., node-layer tuples), each of which corresponds to a specific (node, time) pair. Hence, this structure makes it possible to detect temporally evolving communities in a natural way.

For our computations of community structure, we flatten the $N \times N \times m$ adjacency tensor into an $(Nm) \times (Nm)$ adjacency matrix (known as a 'supra-adjacency matrix' [5,39]), such that the intralayer connections are on the main block diagonal and the interlayer connections occur in the off-block-diagonal entries. We detect communities by maximizing modularity, which we use to describe the 'quality' of a particular network partition into communities in terms of its departure from a null model [14]. The null model amounts to a prior belief regarding influences on network structure, so it is important to carefully consider the choice of null model [19,21,28].

For a weighted static network W, modularity is [40]

$$Q = \frac{1}{2w} \sum_{ij} (W_{ij} - \gamma P_{ij}) \delta(c_i, c_j), \tag{2.1}$$

where $2w = \sum_{ij} W_{ij}$ is the total edge weight, c_i denotes the community that contains node i, the Kronecker delta δ is 1 if $c_i = c_j$ and 0 if $c_i \neq c_j$, and P_{ij} is the ijth element of the null-model matrix. One can examine different scales of community structure by incorporating a resolution parameter γ [41,42]. Smaller values of γ tend to yield larger communities and vice versa.

For multislice networks, modularity is given by

$$\bar{Q} = \frac{1}{2\bar{w}} \sum_{ijsr} [(\bar{W}_{ijs} - \gamma \bar{P}_{ijs}) \delta_{sr} + \delta_{ij} C_{jsr}] \delta(\bar{c}_{is}, \bar{c}_{jr}), \tag{2.2}$$

where $2\bar{w} = \sum_{ijs} \bar{W}_{ijs}$, the quantity $\bar{c}_{is}$ denotes the community that contains node i in layer s, and $\bar{P}_{ijs}$ is the ijth element of the null-model tensor in layer s [25].

To detect communities using modularity maximization, one searches the space of possible network partitions for the one with the highest value of modularity. Because exhaustive search over all possible partitions is computationally intractable [43], practical algorithms use approximate optimization methods (e.g., greedy algorithms, simulated annealing, spectral optimization and so on), and different approaches offer different balances between speed and accuracy [8,9].

In the present paper, we maximize modularity using a two-phase iterative procedure similar to the Louvain method [44]. (See [45] for the code.) We work with the modularity matrix B with elements $B_{ij} = W_{ij} - \gamma P_{ij}$ for static networks and with the modularity tensor with elements $\bar{B}_{ijs} = \bar{W}_{ijs} - \gamma \bar{P}_{ijs}$ for multislice networks.

The employed Louvain-like algorithm [45] is stochastic, and a modularity landscape for empirical networks typically includes a very large number of nearly optimal partitions [18]. For each of our numerical experiments, we thus apply the computational heuristic 100 times to obtain a *consensus community structure* [46] by constructing an *association matrix* A^{rep} (where the entries A^{rep}_{ij} represent the fraction of times that nodes i and j are assigned to the same community in the 100 realizations) and then detecting communities in A^{rep} by maximizing modularity using the uniform null model $P^U_{ij} = 2w/[N(N-1)]$ [28] and $\gamma = 1$. Using the uniform null model helps emphasize the most commonly occurring co-classifications (i.e., assignments of a pair of nodes to the same community) in the association matrix.

For multislice networks, we perform community detection and then consensus clustering using the same basic procedure. This yields an assignment of each multilayer node (i.e., node-layer tuple) to a community. We are also sometimes interested in community assignments of the original entities (i.e., a partition of the set of nodes regardless of what layer they are in). For example, we might wish to compare the result of algorithmic community detection to known partitions, such as grouping a 'physical node' (i.e., province) by climate, population, administrative region, and so on. To do this, we perform what we call *province-level community detection*, which proceeds in two stages: (1) we detect communities in a multislice network using any desired method (and any desired null model); and (2) we then use this partition to construct an $N \times N$ province-level association matrix (i.e., a matrix A^{province} whose entries A_{ij}^{province} represent the fraction of layers in which physical nodes i and j are classified together), and we detect province-level communities by maximizing modularity on this association matrix using a uniform null model. The goal of the optimization in stage (2) is to try to obtain the most temporally persistent community structure in the association matrix (i.e., one that is often detected in multiple layers). We then use consensus community detection across 100 realizations of province-level community structures.

2.1 *Null models for community detection*

The choice of null model is vital for the detection of communities using modularity maximization [18, 19,28]. The most common choice is the NG null model, which randomizes a network such that the expected strength sequence of nodes is preserved [15,47]. For static weighted networks, the NG null model is

$$P_{ij}^{\text{NG}} = \frac{k_i k_j}{2w}, \tag{2.3}$$

where $k_i = \sum_j W_{ij}$ is the strength (i.e., weighted degree) of node i and $2w = \sum_{ij} W_{ij}$ is the total edge weight in the network.

For multislice networks, the NG null model is [25]

$$\bar{P}_{ijs}^{\text{NG}} = \frac{\bar{k}_{is}\bar{k}_{js}}{2\bar{w}}, \tag{2.4}$$

where $\bar{k}_{is} = \sum_j \bar{W}_{ijs}$ is the intralayer strength of node i in layer s and $2\bar{w} = \sum_{ijs} \bar{W}_{ijs}$.

Despite its popularity and demonstrated effectiveness in many situations, the NG null model is naive in the sense that it only takes node strengths into account. It does not incorporate problem-specific information (such as spatial embeddedness). It is often important to incorporate additional (domain-specific or even problem-specific) information, and what one considers to be connected 'at random' depends fundamentally on the research question of interest. Consequently, the NG null model is not always suitable.

2.1.1 *Spatial null models: gravity model.*

In many spatially embedded networks, proximity has a strong effect on the connections between nodes, as (all else held equal) neighbouring nodes are more likely to be connected to each other (and their connections are likely to have larger weights) than nodes that are far away [7,21]. Moreover, proximity can mask other underlying influences. Consequently, incorporating the expected influence of proximity on edge weights into null models for community detection should make it possible to discover new and important types of structures.

Expert *et al.* [21] proposed a spatial null model that was inspired by the 'gravity model' of human mobility [48–52]. A gravity model assumes that the interaction between two locations is proportional to their importance (e.g., population) and that it decays with distance.

In the standard gravity model, the interaction between locations i and j (with respective populations n_i and n_j) that are a distance d_{ij} apart is

$$G_{ij} = n_i^{\alpha} n_j^{\beta} f(d_{ij}), \tag{2.5}$$

where the 'deterrence function' $f(d)$ describes the effect of space on node interactions. Common choices for the deterrence function include inverse proportionality to distance (i.e., $f(d_{ij}) = 1/d_{ij}$), inverse proportionality to squared distance (i.e., $f(d_{ij}) = 1/d_{ij}^2$), exponential decay (i.e., $f(d_{ij}) = e^{-d_{ij}}$) and other interactions of the form $f(d_{ij}) = d_{ij}^{\kappa}$ [7]. It is common to estimate the parameters α, β and κ using regression. Gravity models have been employed successfully during the past half century to model spatial interactions such as population migration [7,53,54], trade [55] and disease spread [36].

The simplest form of a gravity-like interaction in Equation (2.5), with $\alpha = \beta = 1$ and $\kappa = -1$, was incorporated into a *gravity null model* [21] to give

$$P_{ij}^{\text{grav}} = I_i I_j f(d_{ij}), \tag{2.6}$$

where I_i is the importance of node i. One estimates the deterrence function

$$f(d) = \frac{\sum_{\{k,l \,|\, d_{kl}=d\}} W_{kl}}{\sum_{\{k,l \,|\, d_{kl}=d\}} (I_k I_l)} \tag{2.7}$$

for each distance d for all node pairs that are separated by distance d. Expert *et al.* [21] used the population n_i of province i as their measure of node importance. We briefly experimented with several variations, such as using population density or a logarithm of the population [i.e., $I_i = \log(n_i)$] on the disease data sets, and we observed no significant qualitative differences in community structure. When incorporating population into a null model, we thus adopt the choice of Expert *et al.* Another simple choice for importance is node strength (i.e., $I_i = k_i = \sum_j W_{ij}$), though the null model then becomes very similar to the usual NG null model [21]. If $f(d)$ does not depend on distance, then the gravity null model reduces to the NG null model.

In most data sets, distances arise from a discretization of a continuous set of values, so one needs to bin distance data to ensure that there are enough nodes for each distance bin to construct a meaningful deterrence function $f(d)$ in Equation (2.7). Possible binning methods include binning into equal-distance bins (e.g., every b elements, where b is a parameter) and equal-sized bins (i.e., each bin contains a elements, where a is a parameter). We tested the choice of binning procedure on the benchmark networks, applied the gravity null model to empirical data and observed qualitatively similar partitions. Thus, we only report results for equal-distance bins, which we make sufficiently large to ensure that there are always five or more elements in each bin.

For the benchmark networks, we also discuss the influence of bin size on the similarity of algorithmic partitions to planted community structure. We give the specific bin sizes for spatial benchmark and dengue fever correlation networks in the sections in which we discuss the associated experiments.

Combining Equations (2.6) and (2.7) allows us to write the gravity null model as

$$P_{ij}^{\text{grav}} = I_i I_j \frac{\sum_{\{k,l \,|\, d_{kl}=d_{ij}\}} W_{kl}}{\sum_{\{k,l \,|\, d_{kl}=d_{ij}\}} (I_k I_l)}. \tag{2.8}$$

Expert *et al.* used the null model (2.8) to uncover a linguistic partition of a network of Belgian mobile phone calls into the French and Flemish speaking parts of Belgium. This partition was obscured by geographical communities when using the NG null model [21].

In the present paper, we generalize the gravity null model to a multislice setting by calculating a separate contribution to the gravity null model for each layer s. The resulting multislice gravity null model is

$$\bar{P}^{\text{grav}}_{ijs} = I_i I_j \frac{\sum_{\{k,l \,|\, d_{kl}=d_{ij}\}} \bar{W}_{kls}}{\sum_{\{k,l \,|\, d_{kl}=d_{ij}\}} (I_k I_l)}, \tag{2.9}$$

where we have assumed that the population of each province is constant across time. If one has reliable information about changes in population with time, one can incorporate such information into the null model (2.9) by substituting I_i with an analogous quantity I_{is} that depends both on the node i and on the layer s.

2.1.2 *Spatial null models: radiation model.* Gravity models include multiple parameters that one needs to estimate from data or choose in some other way (e.g., arbitrarily). Moreover, by construction, gravity models are unable to predict different fluxes between locations that are the same distance apart but which have regions with different population densities between them. For example, one would expect a higher flux of infectious disease between two locations that are separated by a space with high population density than between locations that are separated by a space with low population density (because of the higher availability of susceptible hosts in the former case) [56]. By contrast, one would expect a smaller commuting flux between a pair of locations in the latter case. This could occur, for example, if there are numerous nearby jobs, as this would reduce peoples' willingness to commute over longer distances [57].

The radiation model [57] was developed to attempt to address these issues. It was designed for population flows and has subsequently been applied successfully in several situations [58,59]. Because the radiation model is designed to capture human mobility between populations and the long-distance spread of many infectious diseases—including dengue fever—is believed to be due largely to long-distance mobility [60], the radiation model may provide a useful-but-idealized description for the spread of disease across space. In this section, we use it to construct a new spatial null model for modularity maximization that we believe may be well-suited for studying the long-distance spread of dengue fever.

The mean commuting flux predicted by the radiation model for locations i and j with populations n_i and n_j, respectively, is

$$T_{ij} = T_i \frac{n_i n_j}{(n_i + r_{ij})(n_i + n_j + r_{ij})}, \tag{2.10}$$

where r_{ij} is the population between locations i and j, and T_i is the number of commuters in location i. A simple way to calculate r_{ij} is to use the population q_{ij} in the circle of radius d_{ij} centred at i and subtract the total of the populations at the origin and destination; that is, $r_{ij} = q_{ij} - (n_i + n_j)$.

Although the radiation model was introduced relatively recently [57], several modifications to it have already been proposed. These include incorporating a normalization for finite systems [59] and the development of a general framework that includes ideas from the radiation, gravity and intervening-opportunities models [61].

We propose a novel null model for modularity maximization based on the original formulation of the radiation model [57]. We use a similar formulation to Equation (2.8) to incorporate both the expected

distance-dependent flux and the actual network structure. To avoid creating a directed network, we use a symmetrized predicted flux

$$\hat{T}_{ij} = (T_{ij} + T_{ji})/2 \tag{2.11}$$

between nodes i and j.[1]

We thereby construct the *radiation null model*

$$P_{ij}^{\text{rad}} = \hat{T}_{ij} \frac{\sum_{\{k,l \,|\, d_{kl}=d_{ij}\}} W_{kl}}{\sum_{\{k,l \,|\, d_{kl}=d_{ij}\}} \hat{T}_{kl}}. \tag{2.12}$$

In Section 4, we will study community structure in empirical data from several years of dengue fever occurrences in Peru. Because we do not possess detailed data on the commuting patterns in Peru (see the description of our data in Section 4.1), we assume that commuters are distributed uniformly across space. We can then simplify Equation (2.10) by substituting $T_i = T_f n_i$, where T_f is the fraction of the population that commutes. Because the quantity T_f is present in both the numerator and denominator of Equation (2.12), we can now cancel it out. However, if one possesses commuting data, it would be desirable to use it to improve the radiation null model.

We also extend the radiation null model to a multislice setting in an analogous manner to the gravity null model. The multislice radiation null model is

$$\bar{P}_{ijs}^{\text{rad}} = \hat{T}_{ij} \frac{\sum_{\{k,l|d_{kl}=d_{ij}\}} \bar{W}_{kls}}{\sum_{\{k,l|d_{kl}=d_{ij}\}} \hat{T}_{kl}}. \tag{2.13}$$

Again, one can incorporate temporal data about population sizes and thereby replace $\hat{T}_{ij}$ with $\hat{T}_{ijs}$ to improve the null model.

2.1.3 *Spatial null models: other models.* The incorporation of spatial information into null models for community detection is an important problem, and several other ideas have been proposed recently. For example, Cerina *et al.* [29] focused on disentangling the correlation between node attributes and space, so they used exponential decay: $f(d_{ij}) = \mathrm{e}^{-d_{ij}/\bar{d}}$, where $\bar{d}$ is the mean distance between nodes in a network. Shakarian *et al.* [31] focused on finding geographically disperse communities, so they introduced a decay constant θ such that $f(d_{ij}) = \mathrm{e}^{-d_{ij}/\theta^2}$. Another recently proposed null model was used to attempt to find geographically proximate communities [30].

As the exact nature of the influence of spatial proximity on interactions in dengue fever transmission is unclear, we choose to focus on null models that include a contribution from the data, rather than using null models with an arbitrarily chosen functional dependence. Thus, we do not test these null models in the present paper.

---

[1] Although the directionality of fluxes is an important factor to study, we wish to keep our null models as simple as possible to help focus instead on the effect of incorporating space into them. Additionally (and again for simplicity), we construct our disease-correlation networks using Pearson correlations (see Section 4.2), and we thereby study the community structure of undirected networks. If one instead constructs a directed network—e.g., by including a time delay when measuring the similarity of time series, by considering ideas such as Granger causality, or by otherwise measuring similarity in a way that produces a directed network (see, e.g., Ref. [62])—then it would also be desirable to construct a directed version of the radiation null model. Clearly, this is an interesting future direction, but it is beyond the scope of our study.

3. Synthetic benchmark networks

To test the performance of the spatial null models, we develop novel synthetic benchmark networks that represent idealized spatially embedded networks with planted community structure.

In what we call the *distance benchmark*, the probability of an edge between two nodes depends only on the geographical distance between nodes and on their community assignments. We assign N nodes uniformly at random to positions on the square lattice $\{1, 2, \ldots, l\} \times \{1, 2, \ldots, l\}$, and we assign a population n_i to each node i (which is an idealized 'city'). We create two versions of the distance benchmark: the 'uniform-population distance benchmark' and the 'random-population distance benchmark'. The uniform-population version corresponds to the benchmark in Expert *et al.* [21]; we assign the same population ($n_i = 100$) to each node. In the random-population benchmark, we assign an integer population uniformly at random from the set $\{1, \ldots, 100\}$.

We plant two equal-sized communities, and we randomly assign each node to one these two communities by selecting nodes uniformly at random without replacement. In the distance benchmarks, the probability p_{ij}^{dist} that an edge is placed between nodes i and j, which are separated by distance d_{ij}, is inversely proportional to the distance:

$$p_{ij}^{\text{dist}} = \frac{\lambda(c_i, c_j)}{Z_1 d_{ij}}, \tag{3.1}$$

where c_i is the community that contains node i, the set c_j is the community that contains node j, and the function $\lambda(c_i, c_j) = 1$ if nodes i and j are in the same community and $\lambda(c_i, c_j) = \lambda_d$ otherwise. The 'inter-community connectivity' λ_d controls the amount of mixing between communities. When $\lambda_d = 0$, only nodes in the same community are adjacent to each other; when $\lambda_d = 1$, there are no distinct communities. The normalization constant Z_1 ensures that $\sum_{i>j} p_{ij}^{\text{dist}} = 1$. There are a total of $L = \mu N(N-1)/2$ edges. We consider edges one by one, and we place a given edge between nodes i and j with a probability of p_{ij}^{dist}. Thus, it is possible for multiple edges, which we interpret as weights, to occur between a given pair of nodes. The parameter $\mu \geqslant 0$ determines the network's edge density. We normalize the edge weights in the network to $[0, 1]$ by dividing each edge weight by the maximum edge weight in the network.

In what we call a *flux benchmark*, we aim to mimic the spread of disease on a network. We allocate edge weights depending on the mean flux between pairs of nodes that is predicted by the radiation model. We place N nodes uniformly at random on the square lattice $\{1, 2, \ldots, l\} \times \{1, 2, \ldots, l\}$, and we assign populations and communities in the same manner as for the distance benchmark. As with the distance benchmark, we consider both uniform-population and random-population versions of the flux benchmark. Now, however, the edge-placement probability p_{ij}^{flux} is directly proportional to the mean predicted radiation-model flux $\hat{T}_{ij}$ (which in turn is inversely proportional to the distance d_{ij}) between nodes i and j. This yields

$$p_{ij}^{\text{flux}} = \frac{\lambda(c_i, c_j)\hat{T}_{ij}}{Z_2}, \tag{3.2}$$

where Z_2 is a normalization constant that ensures that $\sum_{i>j} p_{ij}^{\text{flux}} = 1$.

In Table 1, we summarize the four types of synthetic benchmark networks that we have just introduced.

Using our distance-based and flux-based constructions, we create both static (i.e., single-layer) and multilayer benchmark networks. The static benchmarks enable us to study the performance of modularity maximization using any chosen null model in a simple setting without the additional complications of a multilayer network. However, the multilayer benchmarks are ultimately more appropriate for disease data because they can incorporate temporal evolution.

TABLE 1 *Primary characteristics* (*i.e.*, *population and edge probability*) *for the distance and flux benchmarks for static networks. The notation* rand($\{a, \ldots, b\}$) *signifies that we select a number uniformly at random from the set* $\{a, a+1, \ldots, b\}$. *Additionally,* $\lambda(c_i, c_j) = 1$ *if nodes* c_i *and* c_j *are in the same community and* $\lambda(c_i, c_j) = \lambda_d$ *if they are in different communities,* d_{ij} *is the distance between nodes* i *and* j *in space, and* Z_1 *and* Z_2 *are normalization constants*

Benchmark	Population	p_{ij}
Distance, uniform population	100	$p_{ij}^{\text{dist}} = \dfrac{\lambda(c_i, c_j)}{Z_1 d_{ij}}$
Distance, random population	rand($\{1, \ldots, 100\}$)	$p_{ij}^{\text{dist}} = \dfrac{\lambda(c_i, c_j)}{Z_1 d_{ij}}$
Flux, uniform population	100	$p_{ij}^{\text{flux}} = \dfrac{\lambda(c_i, c_j)\hat{T}_{ij}}{Z_2}$
Flux, random population	rand($\{1, \ldots, 100\}$)	$p_{ij}^{\text{flux}} = \dfrac{\lambda(c_i, c_j)\hat{T}_{ij}}{Z_2}$

The first step in generating multilayer benchmarks involves placing nodes in space and assigning populations in the same manner as for the static benchmarks. We then assign nodes uniformly at random into one of two communities, and we extend this structure into a multilayer planted community structure with m layers. For the 'temporally stable' multilayer benchmarks, the planted community structure is the same for each layer. For the 'temporally evolving' multilayer benchmarks, we change the community assignment of a fraction p of the nodes. For each of these nodes, we select a layer uniformly at random, we assign the node in that layer to a community uniformly at random, and we also change the community of the node in each subsequent layer to that community. (The community assignments for that node in the prior layers are untouched.)

We then generate the edges for each layer independently in the same manner as we generate a static benchmark and using identical parameter values (N, l, μ, λ_d) for each layer (see Fig. 1). Independent generation of each layer using the same parameter values represents differences between observations due to noise and experimental variation.

For each of the above types of multilayer benchmarks, we set the (uniform) weight of the interlayer edges between corresponding nodes in consecutive layers to be $C_{jrs} = \omega \in [0, \infty)$. Because we normalize the intralayer edge weights by the maximum weight (separately in each layer), we obtain synthetic multilayer benchmark networks in which the relative magnitudes of interlayer edges and intralayer edges are comparable with those in the disease-correlation networks.

We report community-detection results that we obtain using the following pipeline. First, we consider $r_1 = 50$ independent instances of a benchmark using the same values of the parameters (N, l, μ, λ_d), (γ, ω), and (when relevant) p. For each of these instances, we perform community detection $r_2 = 50$ times to obtain a consensus community structure. For each of the r_1 consensus community structures, we calculate a summary statistic [such as normalized mutual information (NMI), which we define in Appendix A and briefly discuss soon], and we report the mean of these values.

We evaluate the performance of the NG, gravity and radiation null models on our benchmarks by comparing algorithmic partitions with the planted community structure using NMI [63]. NMI is an information-theoretic similarity measure (see Appendix A for its definition) that is relatively sensitive

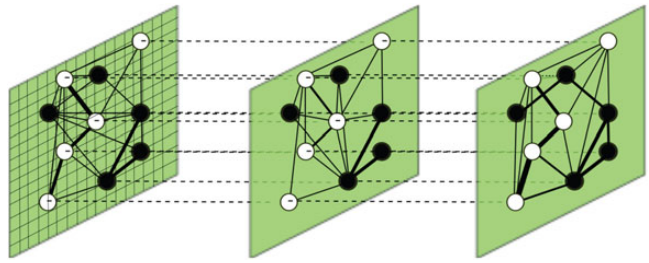

FIG. 1. Construction of a temporally stable multilayer spatial benchmark network. We assign N nodes uniformly at random to positions on an $l \times l$ square lattice (which we show in layer 1) and partition them into two equal-sized communities (black and white), whose constituent nodes we choose uniformly at random. Node i has a population of n_i, and each layer has the same set of nodes. For each layer, we allocate edges uniformly at random according to a probability distribution that depends on the type of benchmark. (For details, see the text and Table 1.) We interpret multiple edges as weights, and we visualize these weights using edge thickness. We connect copies of nodes in adjacent layers using interlayer edges of uniform weight ω (dashed lines).

to small differences in partitions, such as the move of a single node from one community to another, compared with pair-counting measures such as the Rand coefficient and z-Rand scores [64]. This sensitivity makes it suitable for assessing performance on benchmarks that are based on well-defined, ground-truth planted partitions. Additionally, normalization of NMI to lie in the interval $[0, 1]$ facilitates interpretation and comparisons.

We obtain the same qualitative conclusions as for NMI when we use variation of information (VI) [65], which is a different normalized measure of similarity, and with z-Rand scores (see Section 4.3), which is a diagnostic that allows us to compare the coarse structure of disease communities with manual spatial and temporal partitions. See Appendix A for our comparisons using VI and z-Rand scores.

3.1 *Results on static benchmarks*

To emphasize the difference between the gravity and radiation null models, we take $N = 50$ and $l = 10$ to obtain a relatively densely filled lattice. (See Appendix B for the results for synthetic networks with $N = 10$ and $N = 90$.) We compare our results with these parameter values to results that we obtain using the parameter values $N = 100$ and $l = 100$ (which are the parameter values that were used in [21]). We test different bin sizes in uniformly spaced bins using the parameter values $b \in \{10^{-4}, 10^{-3}, 10^{-2}, 0.1\} \cup \{1, 2, \ldots, 10\}$, $l = 10$ and $b \in \{1, 2, \ldots, 100\}$, $l = 100$. We find that bin size makes a large difference for both the distance and the flux benchmarks: $b = 1$ produces the highest NMI scores (i.e., it has the 'best performance'), and increasing bin width leads to a decrease in performance for both spatial null models (see Fig. 2). This effect is especially pronounced for the gravity null model.

The spatial null models have similar performances at the bin sizes that give the highest NMI for each parameter regime for both $l = 10$ and $l = 100$, so we henceforth use the $l = 10$ benchmark with $b = 1$ to use less memory and computational time. We stress, however, that one needs to keep the strong influence of bin size on algorithm results in mind.

We also study the performance of the three null models for modularity maximization for several values of the resolution parameter $\gamma \in \{0.5, 0.75, 1, 1.25, 1.5\}$ and inter-community connectivity $\lambda_d \in \{0, 0.01, \ldots, 0.99, 1\}$ on static benchmark networks with $N = 50$ nodes and lattice-size parameter $l = 10$. Smaller values of γ tend to yield larger communities and vice versa. Considering larger λ_d increases the amount of mixing between the communities and makes it more difficult to successfully detect planted communities. For simplicity, we fix the edge-density parameter to be $\mu = 100$. As we

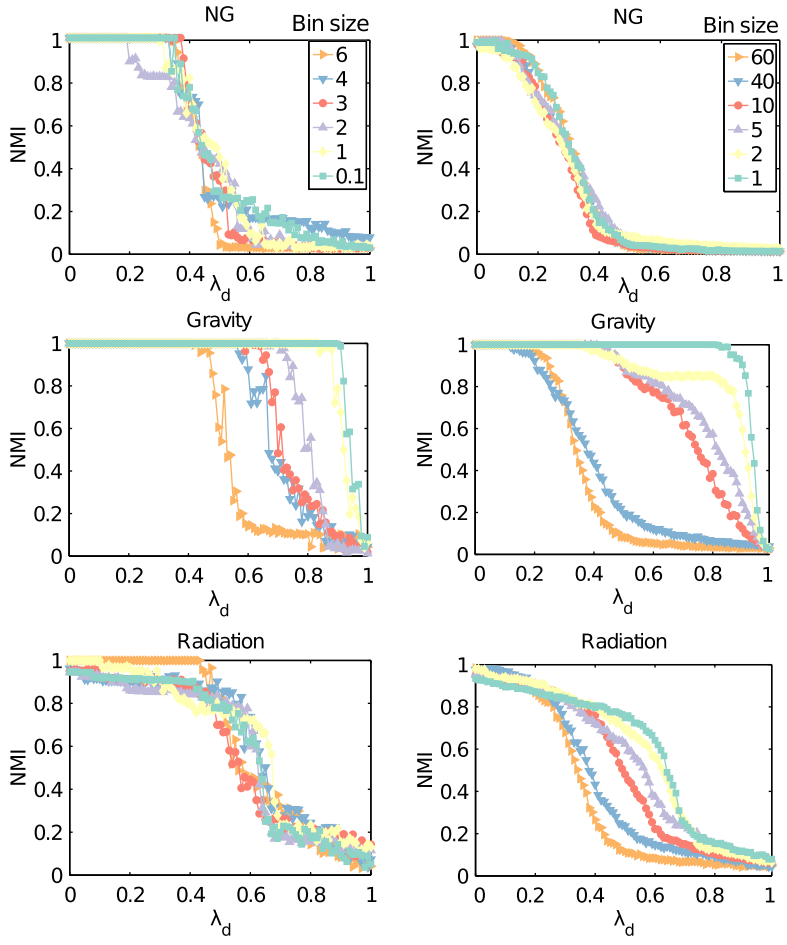

FIG. 2. Uniform-population static benchmarks: normalized mutual information (NMI) scores between algorithmically detected and planted community structures in static uniform-population distance benchmarks for (left) $l = 10$, $N = 50$ and (right) $l = 100$, $N = 100$ with edge-density parameter $\mu = 100$, uniform populations of 100 and different bin sizes (coloured curves). For each plot, the horizontal axis is the inter-community connectivity $\lambda_d \in [0, 1]$. We detect communities by maximizing modularity using the (top) NG, (middle) gravity and (bottom) radiation null models.

discuss in Appendix C, the value of μ has little effect on the results of community detection when it is above a certain minimum.

For the uniform-population distance benchmark, the only factor that influences edge placement is the distance between nodes. For this benchmark, the gravity null model has the best performance; it is able to find the correct partitions for $\lambda_d \lessapprox 0.82$ (see Fig. 3). The radiation null model has the second best performance and is able to find partitions that are broadly similar to the planted partitions for $\lambda_d \lessapprox 0.74$, above which we observe a plateau of 'near-singleton' partitions in which most nodes are placed into singleton communities. (We use the term 'singleton partition' to refer to a partition in which every node is assigned to its own community.) The NG null model, which does not incorporate spatial information, does much worse than either of the spatial null models; it suffers a sharp decline in performance above $\lambda_d \approx 0.4$.

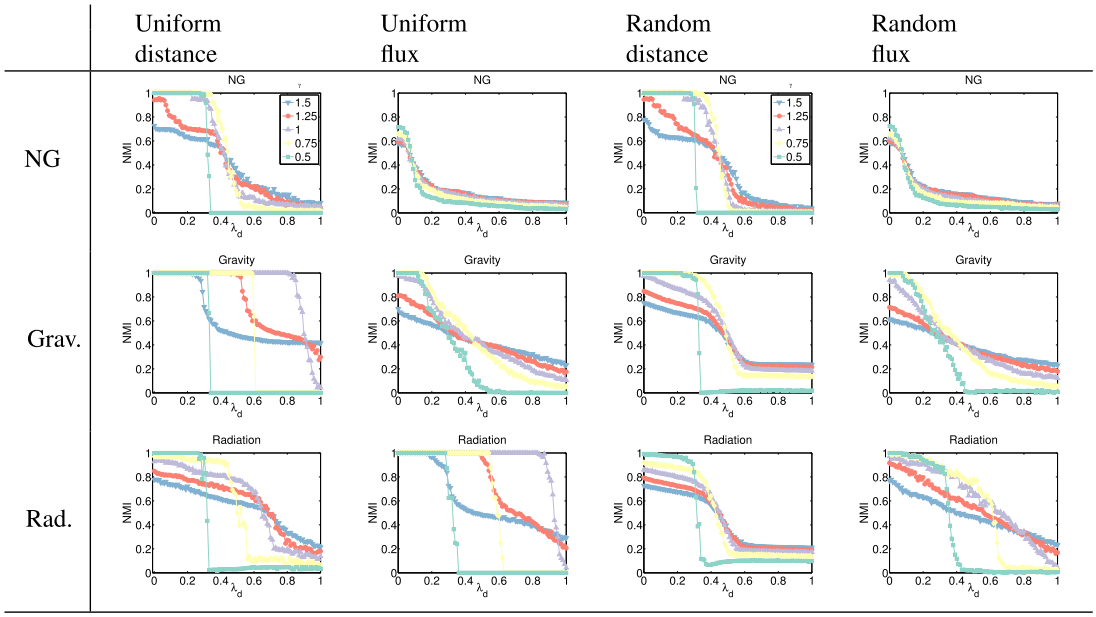

FIG. 3. Static benchmarks: NMI scores between algorithmically detected and planted community structures in static benchmarks with $l = 10$, $N = 50$, $\mu = 100$ and (columns 1 and 2) uniform populations of $n_i = 100$ and (columns 3 and 4) populations n_i determined uniformly at random from the set $\{1, \ldots, 100\}$. We plot NMI for different values of the resolution parameter γ (coloured curves) as a function of inter-community connectivity $\lambda_d \in [0, 1]$. We examine both distance benchmarks (in columns 1 and 3) and flux benchmarks (in columns 2 and 4). We detect communities by maximizing modularity using the (top) NG, (middle) gravity and (bottom) radiation null models.

For the uniform-population flux benchmark—in which we include the population density in the region between two nodes in the estimation of flux (so the population density influences edge structure)—the radiation null model outperforms the other null models. The gravity null model comes in second place, and the NG null model is a distant third.

For the random-population distance benchmark, we observe a fast deterioration in quality of the detected communities for $\lambda_d \gtrapprox 0.4$ for all null models, and all null models reach a 'near-singleton' regime by $\lambda_d \approx 0.6$. The NG null model has the best performance among the three null models for $\lambda_d \lessapprox 0.43$. For $\lambda_d \gtrapprox 0.43$, the gravity null model has the best performance, although the partitions consist largely of singletons for $\lambda_d \gtrapprox 0.6$.

For the random-population flux benchmark, the radiation null model has the best performance of the three null models. It has the slowest decrease in NMI scores as λ_d increases. The gravity null model has the second-best performance, and the NG null model has trouble even when there is no mixing between the two planted communities (see Fig. 3). However, even the best performance is worse for the random-population benchmarks than it is for the uniform-population benchmarks. Including population information into the edge-placement probability by taking $p_{ij}^{\text{distpop}} = n_i n_j \lambda(c_i, c_j)/(Z_3 d_{ij})$ (which yields a 'distance-and-population benchmark'), where Z_3 is a normalization constant, brings back the advantage for the gravity null model (see Appendix D).

Different values of γ can result in rather different performances from the partitions that we obtain from maximizing modularity. Among the parameter values that we consider ($\gamma \in \{0.5, 0.75, 1, 1.25, 1.5\}$), the value $\gamma = 1$ appears to give the best results (i.e., the largest NMI scores).

In the near-singleton regime, $\gamma = 1.5$ outperforms $\gamma = 1$ slightly in terms of NMI scores (see Fig. 3), although the partition that $\gamma = 1.5$ yields is still rather different from the planted partition. As expected, the value of γ that corresponds to the best performance changes with the number of planted communities. Unsurprisingly, the best-performing value of γ tends to become larger as we consider benchmark networks in which we have planted a larger number of smaller communities.

These results demonstrate that, although incorporating spatial information is interesting and potentially beneficial, using a null model that incorporates population information to study community structure in networks whose structure does not depend on population can decrease the performance of community detection. That is, incorporating spatial information is important, but it needs to be done intelligently.

3.2 *Results on multilayer benchmarks*

We now study the influence of the community-detection parameters γ and ω on the community quality of multilayer benchmark networks.

We first study the performance of the NG, gravity and radiation null models for temporally stable uniform-population benchmarks (see Fig. 4) with parameter values $N = 50$, $l = 10$ and $m = 10$ layers using $\gamma \in \{0.5, 0.75, 1, 1.25, 1.5\}$ and $\omega \in \{10^{-3}, 0.1, 0.25, 0.5, 0.75, 1\}$. We want to compare the influence of the resolution parameter γ on our ability to detect planted community structure with the respective results for static networks. Furthermore, for larger values of ω, we expect it to become more common for a node to be assigned to the same community as its counterparts in other layers. This should hinder the detection of our planted communities. However, we observe very little difference in results (and the differences are barely discernible in some plots) across different values of ω for the temporally stable benchmarks. (As we discuss later, we do observe noticeable differences for other benchmarks.)

In Fig. 4, we show experiments on multilayer benchmarks in which we vary γ and use $\omega = 0.1$. Our results on multilayer benchmarks are consistent with our findings for the static benchmarks. Once again, we find that the choice of γ has a large influence on the quality of the partitions that we obtain using modularity maximization, and (as with our findings for static benchmarks) $\gamma = 1$ seems to yield the best performance (i.e., the highest NMI scores) in most cases. The exception is the high-λ_d regime, in which $\gamma = 1.5$ outperforms it slightly. We also observe similar results for random-population benchmarks (see Appendix E) and for both smaller and larger fixed values of ω.

We then examine the NMI between algorithmic versus planted partitions on temporally stable multilayer benchmarks for fixed $\gamma = 1$ and different values of ω and λ_d. As we show in Fig. 5, we find that the value of ω usually has little effect on our ability to detect the planted communities via modularity maximization on benchmarks with a temporally stable community structure. This suggests that perhaps the small interlayer variation from the independent creation of layers does not allow one to observe the influence of ω on community detection, and it is thus important to examine temporally evolving benchmarks.

We then study the performance of the three null models on temporally evolving uniform-population benchmarks (see Fig. 6) with $N = 50$ nodes, a lattice parameter of $l = 10$, a fraction $p = 0.4$ of nodes that change community during the timeline, and $m = 10$ layers. We show results for $\gamma \in \{0.5, 0.75, 1, 1.25, 1.5\}$ with $\omega = 0.1$ and for $\omega \in \{10^{-3}, 0.1, 0.25, 0.5, 0.75, 1\}$ with $\gamma = 1$; compare Fig. 6 with the left panels of Figs. 4 and 5. For temporally evolving benchmarks, varying ω makes a qualitative difference (in contrast to our results for the temporally stable benchmarks). The community structures for $\omega \lessapprox 0.1$ for the gravity null model and $\omega \lessapprox 0.5$ for the radiation null model are the most similar to the planted partitions. This is consistent with our expectation that algorithmically detected community

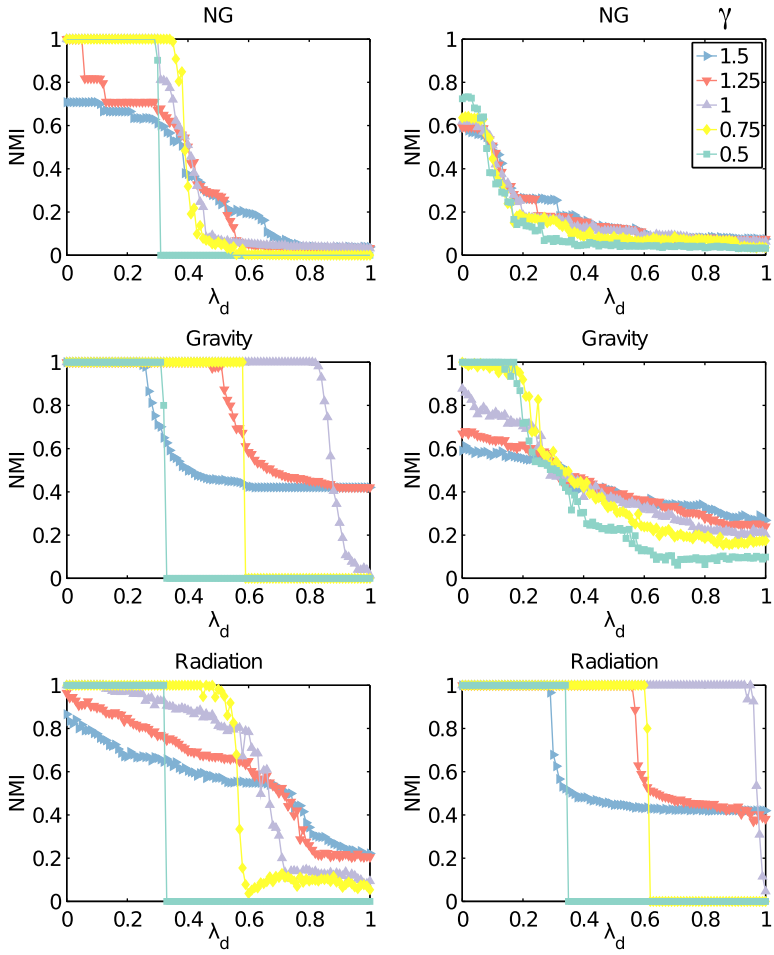

FIG. 4. Uniform-population multilayer, temporally stable benchmarks: NMI between algorithmically detected and planted community structures in uniform-population ($n_i = 100$ for all i) multilayer, temporally stable spatial benchmarks with $N = 50$, $l = 10$, $m = 10$ and $\mu = 100$ for $\omega = 0.1$ and various values of γ (coloured curves) as a function of $\lambda_d \in [0, 1]$ for (left) the distance benchmark and (right) the flux benchmark. We detect communities by maximizing modularity using the (top) NG, (middle) gravity and (bottom) radiation null models.

structure becomes overly biased towards connecting counterparts of nodes across layers above some critical value of ω (that depends on network structure).[2]

We also perform a 'province-level' community detection on the multilayer benchmarks. We seek assignments of nodes (regardless of what layer they are in) to communities and compare these assignments with planted communities. This experiment is meant to model trying to detect spatiotemporal community structure in a disease that persists over time; this is relevant, for example, when one seeks to determine the influence of features like climate on disease patterns. As a simple example, we consider

[2] It is worth comparing our observations with the phase transition in the eigenvalues of a multilayer Laplacian that was reported by Radicchi & Arenas [66] (and has been explored further in subsequent papers).

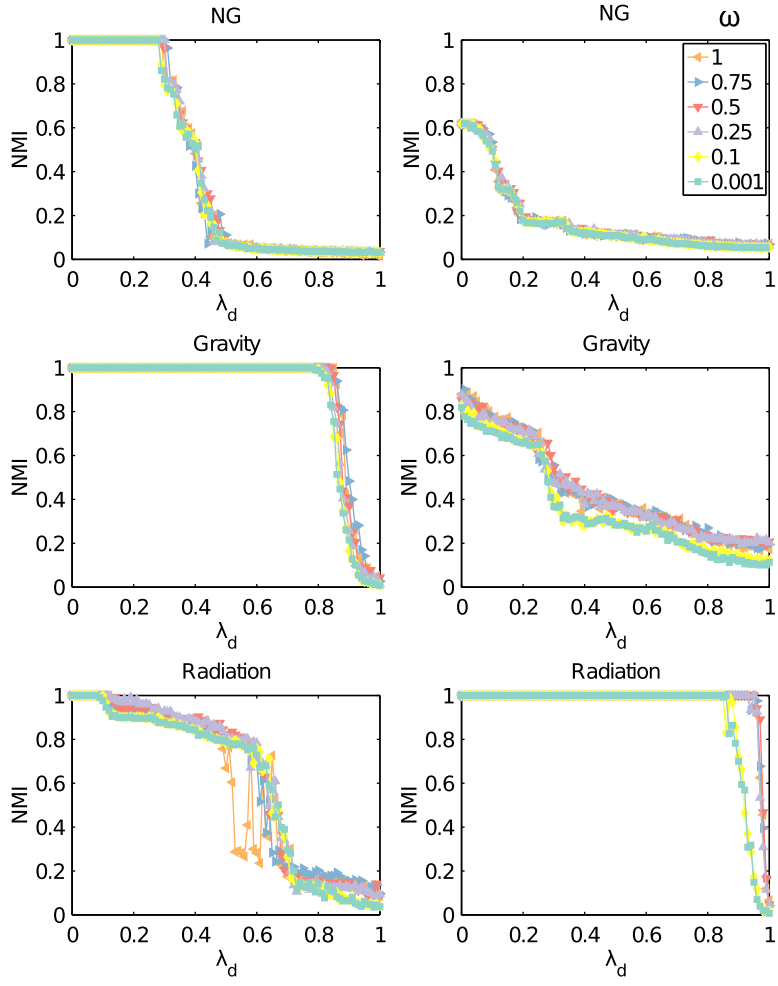

FIG. 5. Uniform-population multilayer, temporally stable benchmarks: NMI between algorithmically detected and planted community structures for uniform-population ($n_i = 100$ for all i) multilayer, temporally stable spatial benchmarks with $N = 50$, $l = 10$, $m = 10$ and $\mu = 100$ for $\gamma = 1$ and different values of interlayer edge weights ω (coloured curves) as a function of $\lambda_d \in [0, 1]$ for (left) the distance benchmark and (right) the flux benchmark. We detect communities by maximizing modularity using the (top) NG, (middle) gravity and (bottom) radiation null models.

the temporally stable multilayer benchmark networks. (One can also consider more stringent benchmarks.) We successfully detect the planted communities, and we obtain similar performance results as with the multilayer communities that we discussed above. (See Appendix F for additional discussion.)

Our results on synthetic benchmark networks suggest that using a spatial null model when detecting communities in a spatial network does not necessarily ensure better performance either in terms of NMI or in more general terms. The quality of results with different null models depends strongly on the network and on the choice of parameter values. In particular, we observe for random-population benchmarks that incorporating population information into the null model causes modularity maximization

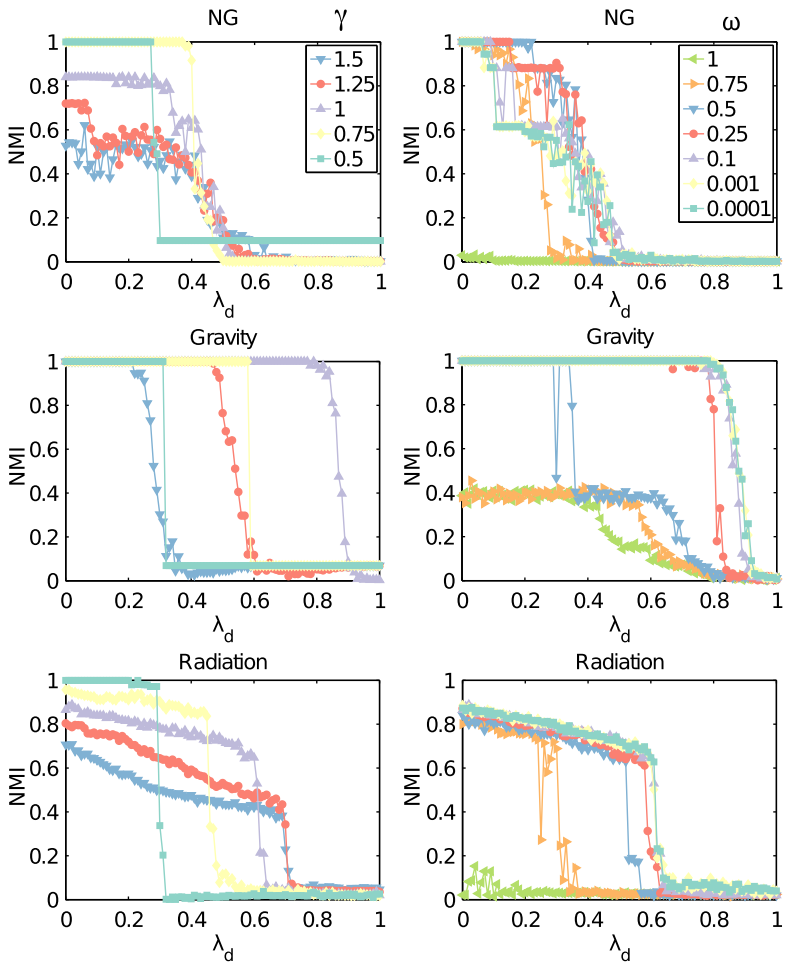

FIG. 6. Uniform-population multilayer, temporally evolving benchmarks: NMI between algorithmically detected and planted community structures in uniform-population ($n_i = 100$ for all i) multilayer, temporally evolving distance benchmarks with $N = 50$, $l = 10$, $m = 10$ and $\mu = 100$ for (left) $\omega = 0.1$ and different values of the resolution parameter γ (coloured curves) and (right) $\gamma = 1$ and different values of the interlayer weights ω (coloured curves) as a function of $\lambda_d \in [0, 1]$. We detect communities by maximizing modularity using the (top) NG, (middle) gravity and (bottom) radiation null models.

to fail to find the planted partitions for a benchmark in which edge placement is independent of node population.

The level of influence of different node properties or events (such as the effect of disease flux on edge placement) and the extent of mixing between communities is unknown for networks that are constructed from real data. For empirical networks, we recommend trying both spatial and non-spatial null models for a wide range of parameter values and studying the results carefully in light of any other known information about the networks. In Section 4, we present an example of using such a procedure to study the community structure of correlation networks that are created from time series of dengue fever cases.

4. Application to disease data

In this section, we assess the performance of the NG, gravity and radiation null models on multilayer correlation networks that we construct from disease incidence data that describe the spatiotemporal spread of dengue fever (a tropical, mosquito-borne viral infection) in Peru from 1994 to 2008.

Disease dynamics are strongly influenced by space, and the distance between regions affects the migration of both humans and mosquitos [56]. Because of the temperature dependence of the mosquito life cycle [67], the dynamics of dengue are also affected by climate. Additionally, different regions of Peru have substantially different climates. Because of such strong spatial effects, it is important to examine and evaluate the performance of different spatial null models when examining communities in networks that are constructed from disease data. We consider dengue fever in the present paper, but this statement holds for disease dynamics more generally.

4.1 *The disease and the data*

Dengue fever (also called simply 'dengue') is a human viral infection that is prevalent in most tropical countries and is carried primarily by the *Aedes aegypti* mosquito [68]. The dengue virus has four strains (DENV-1–DENV-4). Infection with one strain is usually mild or asymptomatic, and it gives immunity to that strain, but subsequent infection with another strain is usually associated with more severe illness [68]. Although dengue was considered to be nearing extinction in the 1970s, increases in human mobility and mosquito abundance have led to its resurgence in many countries—often as recurrent epidemics with an increasing number of cases and severity of disease over time.

Peru is located on the Pacific coast of South America. Its population of about 29 million people is distributed heterogeneously throughout the country. The majority live in the western coastal plain, and there are much smaller population densities in the Andes mountains in the centre and the Amazon jungle in the east. The climate varies from dry along the coast to tropical in the Amazon and cold in the Andes.

Climate heterogeneities influence dengue epidemiology, as increased temperature and humidity increase transmission [69]. The jungle forms a reservoir of endemic disease; from there, the disease occasionally spreads across the country in an epidemic [67]. Additionally, as *Ae. aegypti* typically only travels short distances, human mobility can contribute significantly to the transmission of dengue at all spatial scales [60].

Our dengue data set consists of 15 years of weekly measurements of the number of disease cases across 79 provinces of Peru (which has a total of 195 provinces) and were collected by the Peruvian Ministry of Health [70] between 1994 and 2008. These data were previously analyzed by Chowell *et al.* to study the relationship between the basic reproductive number, disease attack rate, and climate and populations of provinces [67].

Until 1995, the DENV-1 strain dominated Peru; it mostly caused rare and isolated outbreaks [71]. The DENV-2 strain was first observed in 1995–1996 when it caused an isolated large epidemic [72]. DENV-3 and 4 entered Peru in 1999 and led to a countrywide epidemic in 2000–2001 [73], and subsequently there was sustained yearly transmission [71]. The data contains a total of 86,631 dengue cases; most of them are in jungle and coastal provinces (47% and 49%, respectively), and only 4% of the cases occur in the mountains. Disease data was collected in all 195 provinces. A total of 79 provinces reported dengue incidence during the period covered by the data set, but the incidents of disease are never in the data in all 79 provinces at once.

4.2 *Network construction*

Our data set D consists of $N = 79$ time series of weekly disease counts $\{D_1, D_2, \ldots, D_N\}$ for $T = 780$ weeks. The quantity $D_i(t)$ denotes the number of disease cases in province i at time t. See Fig. 7 for a plot of the number of cases versus time. We create networks from these data by calculating the Pearson correlation coefficient between each pair of time series.[3]

We seek to study the temporal evolution of the correlations by constructing separate networks for different time windows. We do two different types of constructions: (1) a sequence of static networks and (2) a multislice network. To create these networks, we divide each of the time series into m time windows by explicitly defining a set of starting times, $\tau = \{\tau_1, \tau_2, \ldots, \tau_m\}$, for each time window. We use Δ to denote the width (i.e., number of time points) of a time window, though the final time window in a multislice network can be shorter than Δ because one can run out of time. Additionally, we use $\tau_1 = 1$.

The starting point τ_t and window width Δ define a time window that we use to select a portion of the disease time series. For example, for the time series of disease cases in province i, the time-series portion $E_i = \{D_i(\tau_t), D_i(\tau_t + 1), \ldots, D_i(\tau_t + \Delta - 1)\}$ represents the numbers of disease cases in province i at times $\tau_t, \tau_t + 1, \ldots, \tau_t + \Delta - 1$. One can use such time series to construct either a sequence of static networks or a multislice network.

To construct a static network, we define a set $\{1, 2, \ldots, N\}$ of N nodes, where node i corresponds to province i. The edge weight

$$W_{ij} = \tfrac{1}{2}(\rho_{ij} + 1) - \delta_{ij} \tag{4.1}$$

represents the similarity between the time series E_i and E_j, and the Kronecker delta δ_{ij} removes self-edges. The quantity ρ_{ij} is the Pearson correlation coefficient between the disease time series (in a given window) for provinces i and j. In a given time window, we let $W_{ij} = 0$ if either node i or node j has no disease cases during that period. For each time window, our construction yields a network W with elements $W_{ij} \in [0, 1]$ that is fully connected among the nodes that experience disease. (If one of these W_{ij} values happens to equal 0, then it is technically almost fully connected.) Identifying each time window according to its starting time, we let $\hat{N}(\tau_t)$ denote the number of provinces in which the disease occurs during the time window with starting time τ_t. When studying static networks, we use the set $\tau = \{1, 2, \ldots, T - \Delta\}$ of starting times to form a set of $T - \Delta$ static networks with overlapping times.

To construct a multislice network, we use the set $\tau = \{1, 1 + \Delta, 1 + 2\Delta, \ldots, 1 + \Delta \times (\lfloor T/\Delta \rfloor - 1)\}$ of starting times to create $\lfloor T/\Delta \rfloor$ non-overlapping time windows. The intralayer edge weights are

$$W_{ijs} = \tfrac{1}{2}(\rho_{ijs} + 1) - \delta_{ij} \tag{4.2}$$

for each layer s. We let $W_{ijs} = 0$ if either node i or node j has no disease cases in layer s. Any interlayer edges associated with a disease-free node in layer s also have weights of 0. If node i has disease cases in a pair of consecutive layers, r and $s \in \{r - 1, r + 1\}$, we connect node i in the rth time window to its counterparts in the contiguous time windows using an interlayer edge of weight $C_{jsr} = \omega \in [0, \infty)$, which is uniform across j, s, and r. (Layer $r = 1$ has interlayer edges only to $r = 2$, the last layer has interlayer edges only to the layer just before it, and all other layers have interlayer edges to both the previous and the next layer.) This yields a weighted multislice correlation network. The case $\omega = 0$ in the multislice network corresponds to a sequence of non-overlapping static networks. See Fig. 7 for a schematic that shows the construction of a multislice network.

[3] Reference [62] compared several methods with calculate similarity networks from time-series data. Our focus in the present paper is on generalizing and evaluating null models, so we use Pearson correlations for simplicity.

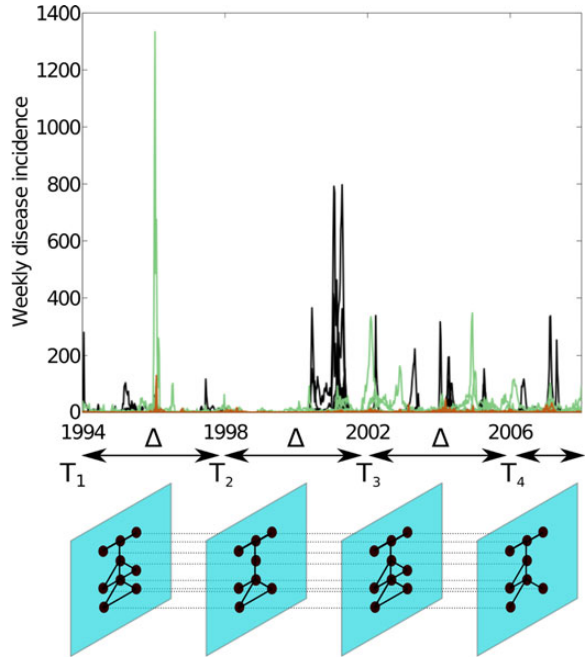

FIG. 7. Construction of a multislice correlation network from disease time-series data. The top panel shows the dengue fever time series for 79 provinces in Peru. We colour the provinces by climate: coastal provinces are in black, mountainous provinces are in brown and jungle provinces are in green. Observe the large epidemics in 1996 (focused in the jungle Utcubamba province) and 2000–2001 (countrywide, but primarily on the northern coast) and the recurrent post-2001 epidemics (which affect various jungle and coastal provinces). The bottom panel shows an example of the multislice network construction for nine nodes with $\tau = \{1, 209, 417, 625\}$ and $\Delta = 208$. (The time points correspond to 1 January 1994, 27 December 1997, 22 December 2001 and 17 December 2005). The nodes represent provinces, and each intralayer edge weight is given by a Pearson correlation between a pair of single-province time series during a given time window. One set of correlations gives one temporal layer, and we connect copies of each node in neighbouring layers using interlayer edges of uniform weight $\omega \in [0, \infty)$ (dotted lines). The case $\omega = 0$ yields a set of non-overlapping static networks.

Similar constructions of (both static and multislice) networks from time series have been employed for systems such as functional brain networks [28,74], currency exchange-rate networks [23] and political voting networks [25,75,76].

Many features, such as the number of layers and the mean and variance of the Pearson correlation values, depend on the parameters that we use in constructing our networks. For example, it is important to consider the choice of the time-window size Δ. There is a trade-off between using sufficiently many layers to obtain a good temporal resolution of events and ensuring that we construct each layer using enough time points to be confident of the statistical significance of the similarity values in the layers of the adjacency tensor [74]. Larger values of Δ yield smaller variations in mean correlation over time and lessen the effects of small, regional epidemics on the number of cases and on the correlation between disease profiles in different provinces. Therefore, we want to use a sufficiently large value of Δ so that we can examine long-term, recurrent disease patterns. Studies based on random matrix theory (RMT) suggest an additional constraint of $\Delta/N > 1$, because correlation matrices generated from time series that are no longer than the number of time series being analysed (i.e., than the number of nodes) are indistinguishable from the correlations that one calculates from short, uncorrelated sequences

of noise [77]. However, choosing a value of Δ that is too large risks over-smoothing data and losing important information.[4] Unless we state otherwise, we use $\Delta = 80$ for the (overlapping) static networks and $\Delta = 60$ for multilayer networks (which never have more than 59 nodes with a positive number of disease cases in any single layer) in order to obtain meaningful correlation matrices while preserving interesting disease patterns. In both cases, each time window contains at least one epidemic season (which occurs yearly during the summer, when the disease is present in the coast and is strengthened in the jungle) to ensure that the networks capture long-term patterns in disease cases.

4.3 *Community structure in disease-correlation networks*

It is well-known that geographical distance has an important influence on disease spread [60,79,80]. Additionally, climate exerts a significant influence on dengue, and it is also necessary to consider Peru's particular topography (as its mountains form a barrier to disease spread) [67,71]. Therefore, we expect community structure in the disease-correlation networks to be strongly geographical when using the NG null model. Applying spatial null models is intended to remove (or at least reduce) the effect of space on community structure, and it may uncover additional effects.

We also expect to observe large changes in community structure at certain time points—such as when the introduction of the new disease strains near 1999 led to large epidemics and the onset of yearly countrywide epidemics [71]. In this section, we explore the similarity of algorithmically detected community structures to spatial and temporal groupings of nodes across a range of parameter values.

To compare the algorithmic partitions of the correlation networks versus manual partitions, we use the z-score of the Rand coefficient (i.e., the so-called 'z-Rand score') [64]. (See Appendix A for the definition of the z-Rand score.) We use z-Rand scores instead of NMI because the former measure is good at detecting similarities in coarse structure [10,64] but is less sensitive to minor changes such as a small number of nodes changing community assignment (see Fig. A1.) For example, suppose that we start with a partition in which provinces are classified by topography into jungle, mountainous and coastal regions. If a small number of jungle nodes later experience different disease patterns than the majority of jungle nodes so that they are now no longer assigned to the jungle community (e.g., they could now be assigned to a second jungle community or to a set of singleton communities), then a z-Rand score that compares the first partition with the new partition would find the two partitions to be rather similar to each other because a large number of the jungle nodes are still assigned to the same community.

For the disease data, we do not possess ground-truth partitions as we did for our synthetic benchmark examples, so we seek to evaluate broad organizational similarities in the algorithmic and manual partitions rather than attempting to conduct a fine-grained evaluation of community structure versus a planted partition. We thereby aim to inform our understanding of structural influences on spatiotemporal patterns of disease spread. One could also aim to search for spatial clustering in community structure without comparing with a manual partition by using, for example, measures of spatial autocorrelation (e.g., Moran's I) [81] or local clustering (e.g., Kuldorff's spatial-scan statistic) [82].

To examine spatial community structures in the static and multilayer networks, we use z-Rand scores to compare our algorithmic partitions to manual partitions. In our 'climate partitions', we group nodes according to the topography of their associated provinces. We start with a 'broad climate partition' that consists of jungle, coastal and mountainous provinces. To obtain a 'detailed climate partition', we then subsequently divide the coastal and mountainous communities into northern, central and southern

[4] See an analogous discussion of time-window choice in Ref. [78] in the context of financial networks.

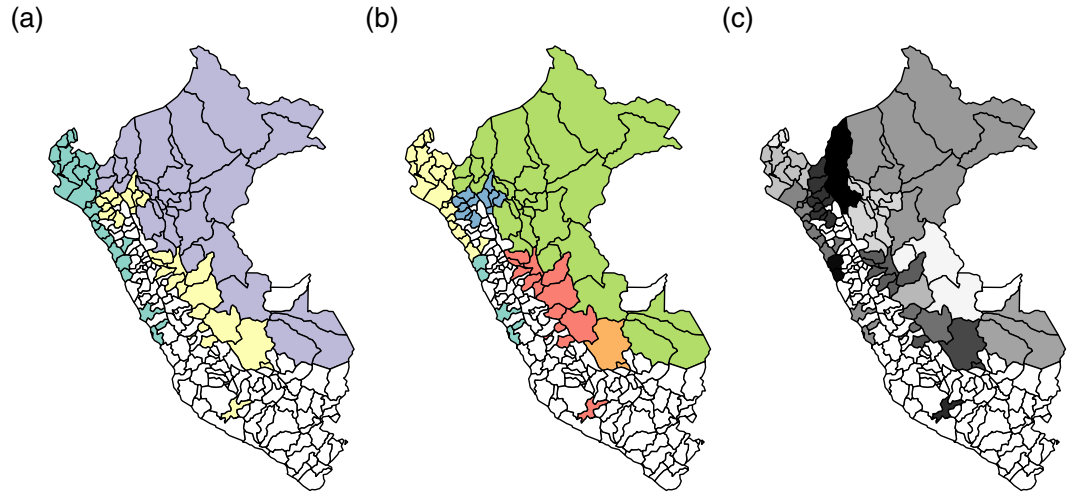

FIG. 8. Visualization of three different spatial partitions of Peru's provinces on a map. (a) Broad climate partition into coast (green; northwest), mountains (yellow; centre) and jungle (purple; northeast); (b) detailed climate partition, in which we start with the broad partition and then further divide the coast and mountains into northern coast, central coast, southern coast, northern mountains, central mountains and southern mountains; and (c) the administrative partition of Peru. We obtained province boundaries from [83] and plot the maps in MATLAB.

provinces [see Fig. 8(a, b)]. In the 19-community 'administrative partition', we assign each node to its associated administrative region [see Fig. 8(c)]. To study spatial features in the disease dynamics, we compare each static network to the detailed climate partition and the administrative partition. For the multilayer networks, we compare each algorithmic partition to a manual partition in which the community assignment of any given node (i.e., province) is the same across all layers.

We use the term 'spatial partitions' to describe partitions that have high z-Rand scores in comparison to the manual climate or administrative partitions. For multilayer networks, we also compare algorithmic partitions to partitions that contain a planted temporal change in community structure. For these comparisons, we group the multilayer nodes into ones that occur before or after a 'critical' time t_c (i.e., partitions into a 'pre-t_c' community and a 'post-t_c' community). We test the set $\tau = \{1, 1+\Delta, 1+2\Delta, \ldots, 1+\Delta \times (\lfloor T/\Delta \rfloor - 1)\}$ of times that we use to create a multilayer network, and we report the time with the highest z-Rand score as the critical time t_c. We also test for pairs of critical times (yielding a partition into three communities) by examining all possible pairs of critical times, t_{c_1} and t_{c_2}, in the same manner. We use the term 'temporal partitions' to describe algorithmic partitions of the disease-correlation networks that yield high z-Rand scores in these comparisons.

4.3.1 *Modularity maximization using the NG null model.* We first study community structure in the 700 overlapping static networks formed by taking $\tau = \{1, 2, \ldots, 700\}$ and using $\Delta = 80$. (There are 779 time points in total.) The community structures that we obtain from maximizing modularity have a strong spatial organization, as suggested by the high z-Rand scores when compared with climate partitions. As one can see in Fig. 9(a), in which we plot the z-Rand scores versus the centres of the time windows that correspond to the static networks, spatial organization is especially evident starting

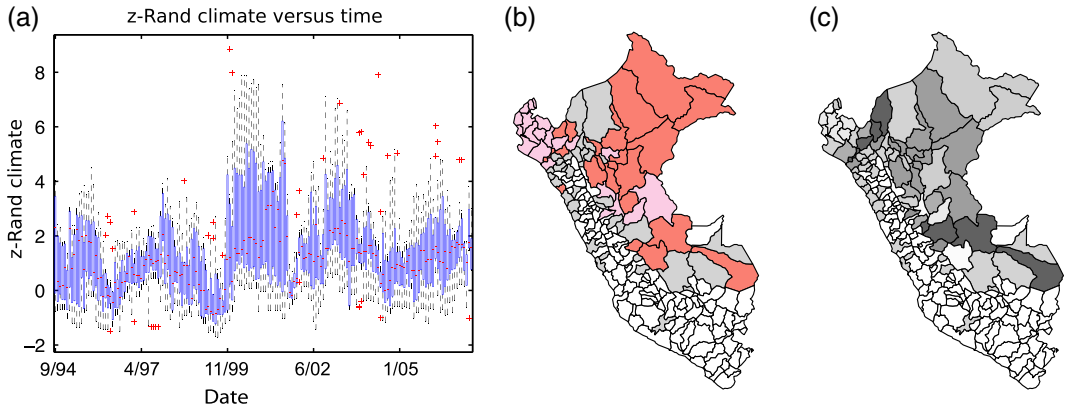

FIG. 9. Properties of algorithmic community structure that we obtain by maximizing modularity using the NG null model for dengue fever static correlation networks that we construct with a time-window size of $\Delta = 80$. (a) A box plot of the z-Rand scores versus the detailed climate partition for $\gamma \in \{0.1, 0.2, \ldots, 2.9, 3\}$ for the 700 static networks that cover the entire time period. (b) Partition with the highest z-Rand score when compared with the detailed climate partition. The resolution-parameter value is $\gamma = 1$, the layer is 293 (which is centred at December 1999), and the z-Rand score is 8.85. (c) Partition with the highest z-Rand score when compared with the administrative partition. The resolution-parameter value is $\gamma = 1.2$, the layer is 453 (which is centred at October 2002), and the z-Rand score is 8.76. In panels (b, c), we use a map of Peru in which we colour provinces according to their community assignment. White provinces are ones in which our data does not include any reported cases of dengue fever in the indicated time window. In panel (a) of the present figure, as well as in our subsequent figures, time points that we indicate on the axes correspond to the centres of the associated time windows.

in the year 2000. In our subsequent figures as well, the time points that we indicate on the axes again correspond to the centres of the associated time windows.

As one can see from a plot of the number of epidemic cases over time (see Fig. 7), this transition seems to occur near the time of the largest countrywide epidemic in the data, and the subsequent period includes recurring yearly epidemics that were linked to climatic patterns in prior studies [71]. There are two periods of significantly spatial partitions: one corresponds to the 2000–2001 epidemic, and it contains the spatial partition that has the highest z-Rand score versus climate [see Fig. 9(b)]; the second occurs in 2002–2004, and it contains the spatial partition that has the highest z-Rand score versus an administrative partition [see Fig. 9(c)]. The z-Rand scores of the spatial partitions decrease starting in 2004 despite the continuing yearly dengue epidemics.

In Fig. 9(b, c), we plot the partitions that have the highest z-Rand scores with respect to the manual climate and administrative partitions. We observe that the high-scoring climate partition includes a large community that is dominated by the jungle (see the red community that includes the northeastern regions) and a large community that is dominated by the coast (see the pink community that includes the northwestern regions), whereas the high-scoring administrative partition is composed of seven small communities. The jungle nodes form the largest communities in each of these spatial partitions. There was a dengue epidemic in most of the provinces in these large communities during the depicted time periods.

We now consider community structure in the multilayer disease network with non-overlapping layers that we construct using the time points $\tau = \{1, 61, \ldots, 721\}$ and a time-window width of $\Delta = 60$. We compare algorithmically computed community structure of the dengue fever multilayer disease-correlation network to manual partitions for several values of the parameters ω and γ in the domain

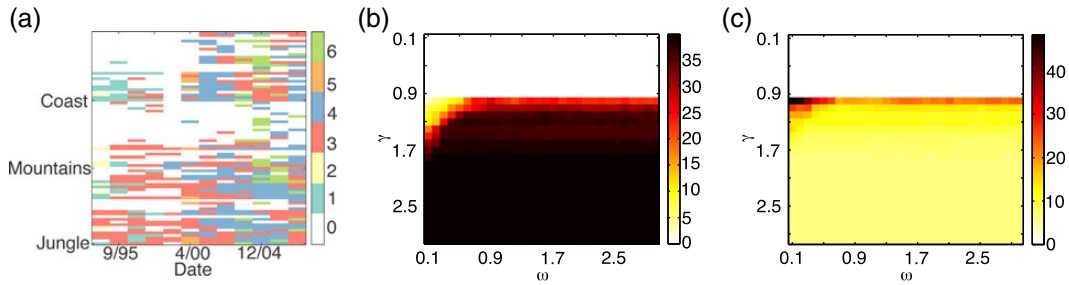

FIG. 10. Algorithmic partitions, which we obtain by maximizing modularity using the NG null model, of the dengue fever multilayer disease-correlation network that we construct using the time-window width $\Delta = 60$. (a) An example of a consensus community structure that we obtain for $\gamma = 1$ and $\omega = 0.1$ across 100 realizations of community detection. We plot layer numbers on the horizontal axis, and we indicate the nodes on the vertical axis. We use colour to represent the community assignment of the nodes. We observe several times at which communities die and new ones are born. (b, c) Results of varying the parameters γ and ω. We show the z-Rand scores for similarity to (b) spatial partitions by climate and (c) temporal partitions before and after a critical time t_c. (For this figure and for each set of parameter values, we select the highest scoring t_c; in the majority of cases, t_c corresponds to January 2002.)

$[0, 3] \times [0, 3]$ (see Fig. 10). For $\gamma \lessapprox 1$, all nodes are in one community. For $\gamma \in [1, 1.2]$ and $\omega \lessapprox 1$, the algorithmically detected partitions have a relatively high z-Rand score when compared with the temporal partition [see Fig. 10(c)]. As we illustrate in Fig. 10(a), the partitions exhibit a mixture of spatial and temporal features.

When studying the qualitative features of the partitions for $\gamma \in [1, 1.2]$ (where the endpoints of this interval are approximate) and $\omega \lessapprox 1$, we observe that multilayer modularity maximization repeatedly finds 2001 as the single critical time, and 2001 and 2005 as the most common pair of critical times t_c (i.e., the strongest change points in temporal community structure). These results agree with visual inspection of Fig. 10(a). These outputs from community detection suggest that a strong shift in the patterns of disease correlations occurred around these times. Indeed, Peru experienced a large countrywide dengue epidemic in 2000–2001, and this period also marks the onset of new yearly epidemics [71]. Thus, our method recovers the most important biological event in this data set in addition to providing additional information about spatial influences on disease spread. We also observe several other times when new communities are born, but we do not know the biological significance of these dates. Notably, in this parameter regime, multilayer modularity maximization using the NG null model does not identify the large epidemic in the jungle Utcubamba province in 1996 (see Fig. 7), which is the other large event in the time period covered by this data set.

The community structure that we detect depends heavily on parameter values. In many parameter regimes (especially when $\gamma \gtrapprox 1$ and $\omega \gtrapprox 0.5$) communities appear to be predominantly spatial, and we find high z-Rand scores when compared with the climate and administrative partitions [see Fig. 10(b)]. The strong influence of spatial proximity on community structure is unsurprising, as geographic distance exerts an important influence on disease spread [60,80]. Previous studies [21,84,85] have also noted that geographic factors strongly influence the partitions that one obtains by maximizing modularity in (static) spatial networks. We also find that the community structures detected by maximizing modularity using a correlation null model that is designed specifically for networks generated by measuring correlations in coupled time series [19] exhibit a similar structure and extent of spatial organization as in the NG null model for both static and multislice networks. This underscores that the disease-correlation network includes strong spatial effects, as even using a null model designed for correlation structure

misses salient features of the data. Indeed, other ingredients that help shape community structure in the multilayer dengue fever correlation network may be obscured by the strong influence of spatial proximity. However, it is conceivable that we may be able to reveal such effects by using a null model that incorporates spatial effects. We pursue this idea in Section 4.3.2.

4.3.2 *Modularity maximization using spatial null models.* In this subsection, we maximize modularity in the dengue fever correlation networks using spatial null models. We obtained province locations from the Geonames.org website [86], and we obtained their populations from the Peruvian Instituto Nacional de Estadística e Informática (INEI) [70]. We were only able to obtain the 1994 and 2007 populations; because of the limited range of data and the several changes in Peruvian administrative structures between 1994 and 2007, our calculations use only the 2007 populations.

The maximum inter-province distance in Peru is about 1300 km. We report numerical experiments using a bin size of 400 km after testing the spatial deterrence for several other sizes (ranging between 50 km and 500 km) in the same manner as in Ref. [21]. That is, we study the shape of the deterrence function [see Equation (2.5) and the nearby discussion] versus distance for different bin sizes, and we then examine the community structures that we obtain using these different bin sizes. Although bin sizes do affect the shape of the deterrence function—smaller sizes tend to give smoother results—all of the bin sizes that we tested produce very similar partitions for both the gravity and radiation spatial null models. We select the smallest bin size that guarantees that there are at least five pairs of nodes in each bin to ensure that the averaging that is necessary for calculating the deterrence function uses at least five elements. From a practical perspective, we recommend testing the bin-size parameter and studying both the shape of the deterrence function and the resulting community structure for several suitable bin sizes. Based on our results for spatial benchmarks in Fig. 2, we expect bin size to influence the results of community detection in situations in which there is an underlying structure that is masked by spatial proximity. As suggested by Expert *et al.* [21], one could select a 'representative' structure by comparing the partitions for different bin sizes using a similarity measure such as NMI.

Recall from Section 4.1 that only 79 of the 195 provinces include reported cases of dengue fever in our data, so we use location and population data only for those provinces.

We first study community structure for static disease-correlation networks by maximizing modularity using the gravity and radiation null models. Both null models seem to remove most of the spatial organization of the community structures (including the temporal variation in the spatial correlations), as indicated by low values and low variation of spatial z-Rand scores (not shown). For both the gravity and radiation null models, we observe a broad similarity in community structure across time for a variety of values of the resolution parameter γ [see Fig. 11(a)]. In particular, these structures consist of one dominant community, which contains the majority of nodes at any given time, and several singleton communities [see Fig. 11(b)]. The singleton communities tend to be the most populous provinces.

We also examine community structure in multilayer correlation networks by maximizing modularity using spatial null models. We again obtain partitions that include one large community with most of the provinces [see Fig. 12(a, b)], although several of the most populous provinces are assigned to communities that consist only of the given province and its counterparts across time. We refer to such a community as a 'single-province community'. This situation occurs for all of the tested parameter values. Additionally, we do not observe any clear pattern in the z-Rand scores across different values of γ and ω.

Our findings suggest that the use of a spatial null model for modularity maximization may remove the majority of the variation in the correlation structure of the dengue fever correlation networks, such

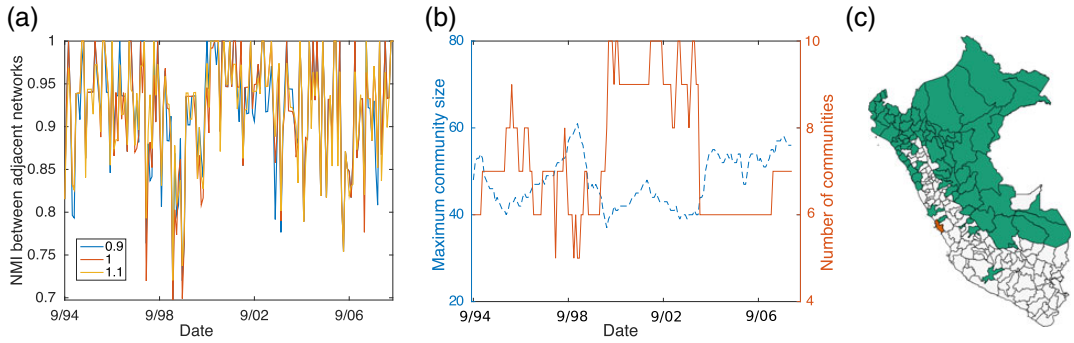

FIG. 11. Properties of algorithmic community structure that we obtain by maximizing modularity using the gravity null model for dengue fever static correlation networks that we construct with a time-window size of $\Delta = 80$. (a) NMI between contiguous layers for $\gamma \in \{0.9, 1, 1.1\}$. (b) Maximum community size (blue dashed curve), which indicates the number of provinces in the largest community, and number of communities (orange solid curve) for $\gamma = 1$. (c) Community structure that has the highest z-Rand score versus climate among the dengue fever static correlation networks that we construct using $\Delta = 80$. (The resolution-parameter value is $\gamma = 2.9$, the layer is centred at January 1996, and the z-Rand score is 4.94.) We show the community structure on a map of Peru, and we colour provinces according to their community assignment. White provinces are ones in which our data does not include any reported cases of dengue fever in the indicated time window. Observe the single giant community that contains almost all of the nodes except the Lima province (which is a singleton that includes 41% of the population).

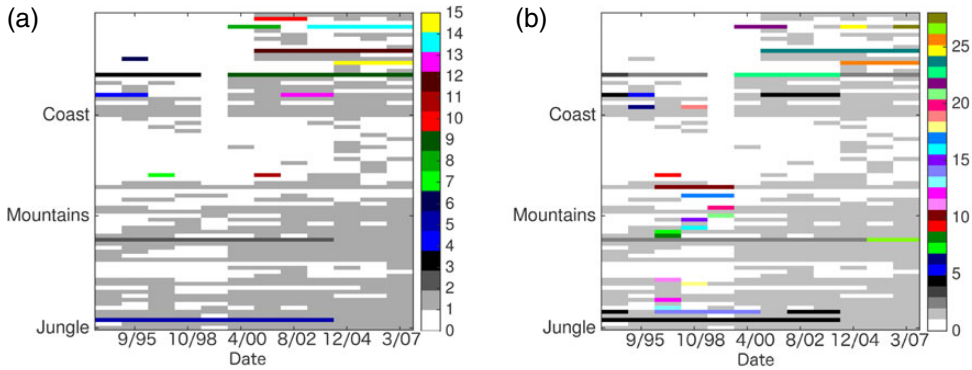

FIG. 12. Consensus community structure that we obtain by maximizing modularity for 100 realizations using (a) the gravity null model and (b) the radiation null model of the dengue fever multilayer disease-correlation network that we construct using a time-window size of $\Delta = 60$. We use a resolution-parameter value of $\gamma = 1$ and an interlayer coupling strength of $\omega = 0.1$. We indicate dates (where each date corresponds to the centre of a time window) on the horizontal axes, and we indicate nodes on the vertical axes. We colour nodes according to their community assignment.

that the influence of population size may be the only major factor that remains. There are only five provinces whose populations are over 500,000, and these provinces are often assigned to singleton communities or to single-province communities when we use a spatial null model. This suggests that they have different disease patterns from the other provinces. One could speculate whether this has any relation to a minimum population size required for sustained disease transmission, which for dengue has been estimated to lie between 10,000 and 500,000 [67,87].

4.3.3 *Province-level multilayer communities.* We further examine relationships between the disease patterns in provinces by examining 'province-level communities' in multilayer dengue fever correlation

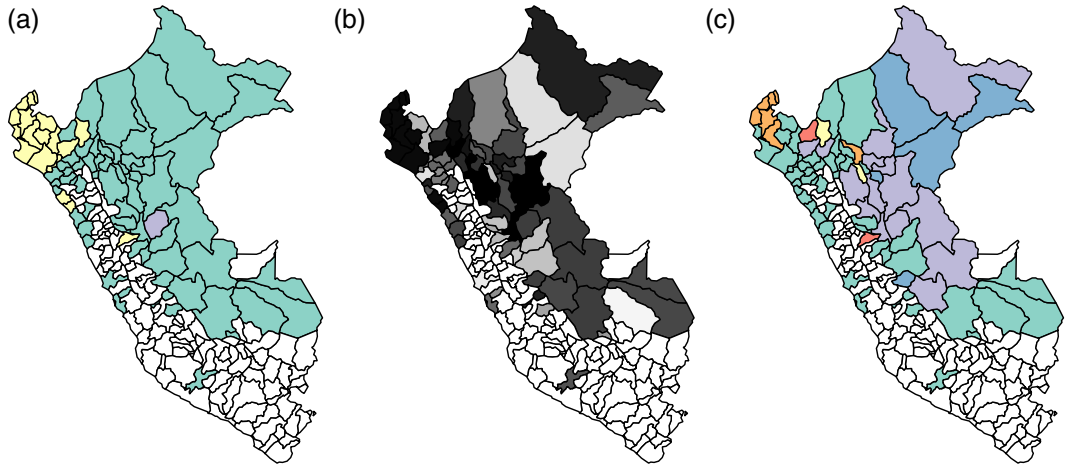

FIG. 13. Province-level algorithmic community structure, which we obtain by maximizing modularity, in static and multilayer dengue fever correlation networks. We colour the provinces according to their community assignment. White provinces are ones in which our data does not include any reported cases of dengue fever in the indicated time window. In panels (a, b), we show the results of using an NG null model in a fully aggregated network (i.e., $\tau = 1$ and $\Delta = 779$) with a resolution-parameter value of (a) $\gamma = 1$ and (b) $\gamma = 1.1$. In panel (c), we show the province-level communities that we obtain using a time-window width of $\Delta = 60$, the NG null model, $\gamma = 1$, and $\omega = 0.1$.

networks. (Recall from Section 2 that 'province-level communities' are sets of physical nodes—i.e., the provinces—rather than node-layer tuples, which describe provinces during a specified time window.) The simplest approach is to construct a single static network from the entire length-T time series and then detect communities in that network. However, our multilayer approach allows us to aggregate data less severely. This, in turn, allows us to lose less information when examining similarities in disease patterns among the provinces.

When we aggregate the time series over the entire time range to construct a single similarity network (i.e., we choose $\tau = 1$ and $\Delta = 779$), the community structures that we obtain via modularity maximization with the spatial null models all consist of a single large community and a few singleton communities (see Fig. G1 in Appendix G). Only the NG null model is able to detect meaningful-looking communities, which we show for $\gamma = 1$ and $\gamma = 1.1$ in Fig. 13(a,b). For $\gamma = 1$, we find three communities; one is a single jungle province, one consists almost exclusively (15 of its 17 nodes) of northern coastal provinces, and the third contains the remainder of the provinces. This partition has a z-Rand score versus climate of about 7.3. For $\gamma = 1.1$, maximizing modularity using the NG null model yields 28 communities, and many of them are small. The nodes that are assigned to the community of northern coastal provinces are still assigned to the same community for $\gamma = 1.1$. These are the provinces of Peru that were most strongly involved in the 2000–2001 dengue epidemic; 15 of the nodes in this community experienced this epidemic, whereas only two other nodes experienced it.

When detecting communities, data aggregation across all time in the coupled time series results in the 2000–2001 epidemic dominating all other events. By detecting communities in a temporarily evolving multilayer network and examining province-level structure, we hope to obtain more interesting partitions (as this will involve milder data aggregation).

We also study the structure of province-level communities that we obtain from modularity maximization using the uniform null model (with $\gamma = 1$ only) on an association matrix A^{province}. As we discussed in Section 2, we create this matrix by tracking how frequently provinces are classified together

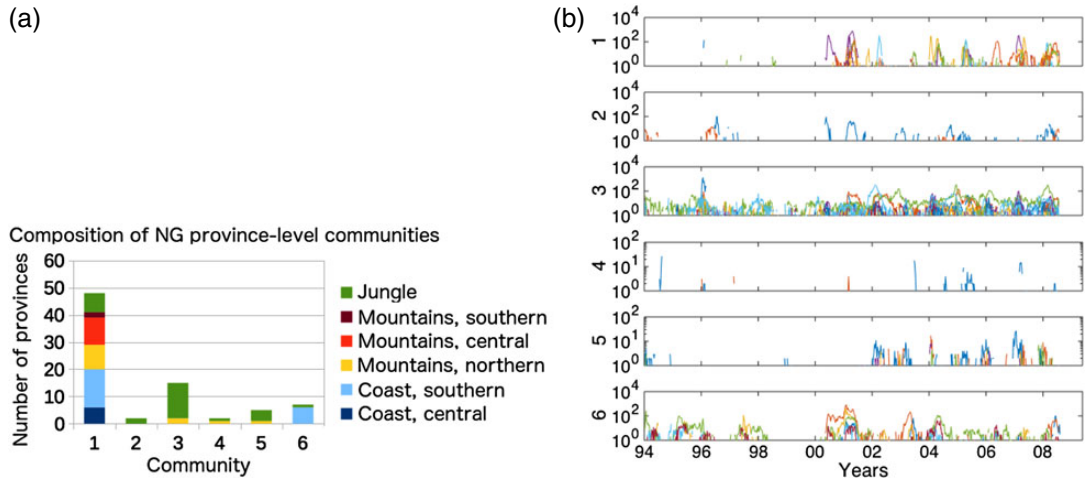

FIG. 14. Structure of the consensus province-level communities that we compute by maximizing modularity in multilayer dengue fever networks for $\gamma = 1$ and $\omega = 0.1$ using the Newman–Girvan null model. (a) Climate composition of the communities. The horizontal axis gives a numerical label for each community. (b) Time series for disease occurrences in the provinces that belong to each of those communities.

(i.e., assigned to the same community) across layers in a multislice network, and we then detect communities in the association matrix. When maximizing multilayer modularity in these experiments, we use the parameter values $\gamma = 1$ and $\omega = 0.1$. We do 100 realizations of the above procedure to obtain 100 different association matrices A^{province}. We detect communities in each of these association matrices, and we form another association matrix based on the fraction of times that the provinces are classified together during this step. We then obtain consensus province-level communities by detecting communities in this second association matrix by again maximizing modularity using the uniform null model (with $\gamma = 1$).

Comparing the province-level communities that we obtain by maximizing modularity using the NG null model to the broad topographical categories of coast, mountain and jungle reveals a large-scale correlation between climate and disease patterns. More than 40 nodes are assigned to one large community that includes central coast, northwestern and southern jungle, and eastern jungle; and coastal north nodes form smaller, strongly spatial communities. [See Fig. 13(c) and Fig. 14.] When we study province-level communities, we observe both distinct types of disease incidence patterns in each community (such as provinces with year-round versus summer infections) and distinct forms of long-term temporal dynamics (such as provinces affected before the large 2000–2001 epidemic and those only affected afterwards).

5. Conclusions

In this paper, we examined time-dependent community structure and the effect of different null models that incorporate spatial information on the results of modularity maximization. We conducted our computational experiments using both novel synthetic benchmark spatial networks and correlation networks constructed from spatiotemporal dengue fever incidence data in provinces of Peru (a situation that is influenced strongly by spatial effects). We compared our results for the standard NG null model versus two null models that incorporate spatial information: a gravity null model [21] and a novel radiation null model.

Our results indicate that it is very important to incorporate problem-specific features, such as spatial information, into the null models for community detection. Our results also illustrate that there are many nuances to consider. That is, it is not simply a matter of incorporating spatial information in an arbitrary way; instead, it is necessary to develop spatial null models that are motivated by application-appropriate generative models. For example, the NG null model performs better than the spatial null models (which both use population data) on the random-population distance benchmark, in which populations vary but edge weight does not depend on them. However, when we remove the variation in population information or modify the benchmark to include population in edge-placement probabilities, we find (as expected) that the gravity null model has the best performance among the null models that we considered.

Parameter choices can also be extremely important, as demonstrated by the strong influence of bin size (when binning distances for the spatial null models) on the results of community detection, our failure to find meaningful communities with any of the null models at low edge densities, and the strong influence of the resolution parameter γ on the results.

To summarize, one needs to consider seriously what features influence the connections in a system, which of those features one wants to try to incorporate into a null model, be careful about including spurious information and test how results change when one changes parameter values. Finally, not incorporating space at all can be more appropriate than incorporating it in a manner that is overly naive. See, for example, our results on the random-population benchmark networks.

In our consideration of dengue fever data, we observed for static networks that maximizing modularity using the NG null model finds structures that are strongly spatial—especially after the onset of yearly epidemics in 2000. In our study, we observed that spatial partitions are often dominated by large communities of neighbouring jungle nodes that experience local epidemics during the time window.

On a multilayer network, maximizing modularity using the NG null model can result in either spatial or temporal partitions (depending on the parameter regime). Temporal partitions successfully find the most important time point in the history of the disease—namely, the introduction of a new disease strain that caused a large epidemic in 2000–2001 and a subsequent change in disease patterns—and several other potentially interesting time points and periods of high spatial correlation.

When studying province-level connectivity, we illustrated that detecting consensus province-level communities in an association matrix that we constructed by examining communities in a multilayer network across time is a better approach than complete data aggregation. When aggregating into a static network, maximizing modularity using any of the tested null models except the NG null model failed to detect any meaningful communities. Modularity maximization using the NG null model was able to reveal the large 2000–2001 epidemic. Aggregating networks results in loss of information, which can be very undesirable when attempting to discern meaningful patterns [3,5]. When we constructed multilayer networks and computed consensus communities, we found ‘spatial’ multilayer partitions and province-level partitions that highlight the importance of climate to the disease patterns of dengue. In these partitions, the jungle provinces are assigned to distinct communities from most mountainous and coastal provinces. This is sensible, as the yearly epidemic patterns tend (on average) to exhibit an earlier epidemic onset in the jungle [67,71], and the jungle climate is rather distinct from the climates in coastal and mountainous provinces. The main climatic difference between jungle provinces and other provinces is temperature, and the influence of temperature on dengue transmission [69,88] and attack rate and persistence has been documented [67,89].

When we attempted to take into account the influence of space by using the gravity and radiation null models, we obtained one large community that contains all but the highest-population provinces (which are assigned to singleton communities for static networks and single-province communities in

multilayer networks). This is different from the Belgian mobile-phone example in Ref. [21], in which maximizing modularity using a spatial null model based on the gravity model revealed a linguistic partition that the authors were not able to detect when using the NG null model. Our results suggest that the incorporation of space into the null model for modularity maximization can sometimes account for the majority of the structure in the disease correlation networks that we studied. The spatial structure that we incorporated directly into the null model likely includes the structure that corresponds to the climatic variation that leads to different epidemic patterns in the jungle, coastal and mountainous provinces. Province population is the only feature that we were able to identify as associated with community structure when using spatial null models: the highly populated (and typically coastal) provinces formed singleton (and single-province) communities. These populous provinces are local economic centres; many people travel there from other provinces and can thereby play a role in disease transmission [60].

The highly populated provinces can also be the seeds of epidemics in other coastal and mountainous provinces, and two studies have in fact reported (so-called) 'hierarchical' transmission of dengue from populous regions to those with low populations in both Peru and Thailand [67,90]. This situation could lead to high positive correlations across distances that are atypically long, which could in turn cause populous provinces to be assigned to singleton (and single-province) communities. Additionally, it is known that population size strongly influences dengue transmission: the basic reproductive number R_0 (i.e., the mean number of secondary cases per infected individual) and disease persistence (as measured by the fraction of weeks that include disease cases) are positively correlated with population size, and mosquito attack rates are negatively correlated with population size [67,71].

The incorporation of spatial information into null models for community detection is both interesting and desirable. As we have illustrated in the present paper, however, there are many nuances that are important to consider. We have also demonstrated that it is important to develop null models that incorporate generative mechanisms for human mobility and flux. We expect that domain-dependent, mechanistic null models will also be crucial in many other applications.

Acknowledgements

We thank Marya Bazzi, Paul Brodersen, Andrew Elliott, Paul Expert, Lucas Jeub, Heather Harrington, Felix Reed-Tsochas, Jesús San Martín and several anonymous referees for helpful comments. We thank Vittoria Colizza and Chiara Poletto for several extensive discussions and comments on the manuscript, and we thank Andrew Elliott for reading and commenting on this manuscript and for his help with creating the figures. Additionally, we thank Yulian Ng and Tom Prescott for their contributions at early stages of this project. The dengue fever data were collected by the Peruvian Ministry of Health [91]. We obtained province locations from the Geonames.org website [86], and we obtained their populations from the Peruvian Instituto Nacional de Estadística e Informática (INEI) [70].

Funding

M.S. was supported by the EPSRC Systems Biology Doctoral Training Centre, the James S. McDonnell Foundation (grant 220020177) and the European Commission FET-Proactive project PLEXMATH (grant 317614). G.C. acknowledges financial support from grant 1R01GM100471-01 from the National Institute of General Medical Sciences (NIGMS) at the United States National Institutes of Health; the Multinational Influenza Seasonal Mortality Study (MISMS) led by the John E. Fogarty International Center, United States National Institutes of Health [92]; the Division of International Epidemiology and Population Studies, the John E. Fogarty International Center, United States National Institutes of Health;

NSF grant 1414374 as part of the joint NSF–NIH–USDA Ecology and Evolution of Infectious Diseases program; United Kingdom Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/M008894/1; and NSF–IIS grant 1518939. M.A.P. was supported by the European Commission FET-Proactive project PLEXMATH (grant 317614), a grant (EP/J001759/1) from the EPSRC and the James S. McDonnell Foundation (grant 220020177).

References

1. Newman, M. E. J. (2010) *Networks: An Introduction*. Oxford: Oxford University Press.
2. Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
3. Holme, P. & Saramäki, J. (2012) Temporal networks. *Phys. Rep.*, **519**, 97–125.
4. Holme, P. & Saramäki, J. (2013) *Temporal Networks*. Berlin, Heidelberg, New York: Springer.
5. Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. & Porter, M. A. (2014) Multilayer networks. *J. Complex Networks*, **2**, 203–271.
6. Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z. & Zanin, M. (2014) The structure and dynamics of multilayer networks. *Phys. Rep.*, **544**, 1–122.
7. Barthélemy., M. (2011) Spatial networks. *Phys. Rep.*, **499**, 1–101.
8. Fortunato, S. (2010) Community detection in graphs. *Phys. Rep.*, **486**, 75–174.
9. Porter, M. A., Onnela, J.-P. & Mucha., P. J. (2009) Communities in networks. *Notices Amer. Math. Soc.*, **56**, 1082–1097, 1164–1166.
10. Traud, A. L., Mucha, P. J. & Porter, M. A. (2012) Social structure of Facebook networks. *Physica A*, **391**, 4165–4180.
11. Waugh, A. S., Pei, L., Fowler, J. H., Mucha, P. J. & Porter, M. A. (2012) Party polarization in Congress: a network science approach. arXiv:0907.3509.
12. Lewis, A. C. F., Jones, N. S., Porter, M. A. & Deane, C. M. (2010) The function of communities in protein interaction networks at multiple scales. *BMC Syst. Biol.*, **4**, 100.
13. Spirin, V. & Mirny, L. A. (2003) Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. USA*, **100**, 12123–12128.
14. Newman, M. & Girvan, M. (2004) Finding and evaluating community structure in networks. *Phys. Rev. E*, **69**, 026113.
15. Newman, M. E. J. (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, **74**, 036104.
16. Lambiotte, R., Delvenne, J.-C. & Barahona, M. (2009) Laplacian dynamics and multiscale modular structure in networks. arXiv:0812.1770.
17. Lambiotte, R., Delvenne, J.-C. & Barahona, M. (2015) Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE Trans. Net. Sci. Eng.*, **1**, 76–90.
18. Good, B. H., de Montjoye, Y. A. & Clauset, A. (2010) Performance of modularity maximization in practical contexts. *Phys. Rev. E*, **81**, 046106.
19. MacMahon, M. & Garlaschelli, D. (2015) Community detection for correlation matrices. *Phys. Rev. X*, **5**, 021006.
20. Bazzi, M., Porter, M. A., McDonald, M., Fenn, D. J., Williams, S. & Howison, S. D. (2014) Community structure in multilayer networks of financial-asset correlations. *Multiscale Model. Simul.*, in press. arXiv:1501.00040.
21. Expert, P., Evans, T. S., Blondel, V. D. & Lambiotte, R. (2011) Uncovering space-independent communities in spatial networks. *Proc. Natl. Acad. Sci. USA*, **108**, 7663–7668.
22. Berger-Wolf, T. Y. & Saia, J. (2006) A framework for analysis of dynamic social networks. *KDD'06: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 523–528.

23. Fenn, D. J., Porter, M. A., McDonald, M., Williams, S., Johnson, N. F. & Jones, N. S. (2009) Dynamic communities in multichannel data: an application to the foreign exchange market during the 2007–2008 credit crisis. *Chaos*, **19**, 033119.

24. Chan, S.-Y., Hui, P. & Xu, K. (2009) Community detection of time-varying mobile social networks. *Complex Sciences* (J. Zhou, ed.). Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 4. Berlin, Heidelberg: Springer, pp. 1154–1159.

25. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J.-P. (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science*, **328**, 876–878.

26. Kawadia, V. & Sreenivasan, S. (2012) Sequential detection of temporal communities by estrangement confinement. *Sci. Rep.*, **2**, 794.

27. Chen, Y., Kawadia, V. & Urgaonkar, R. (2013) Detecting overlapping temporal community structure in time-evolving networks. arXiv:1303.7226.

28. Bassett, D. S., Porter, M. A., Wymbs, N. F., Grafton, S. T., Carlson, J. M. & Mucha, P. J. (2013) Robust detection of dynamic community structure in networks. *Chaos*, **23**, 013142.

29. Cerina, F., De Leo, V., Barthelemy, M. & Chessa, A. (2012) Spatial correlations in attribute communities. *PLoS One*, **7**, e37507.

30. Hannigan, J., Hernandez, G., Medina, R. M., Roos, P. & Shakarian, P. (2013) Mining for spatially-near communities in geo-located social networks. Association for the Advancement of Artificial Intelligence - Social Networks and Social Contagion: Web Analytics and Computational Social Science, Arlington, VA. November 15–17.

31. Shakarian, P., Roos, P., Callahan, D. & Kirk, C. (2013) Mining for geographically disperse communities in social networks by leveraging distance modularity. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM: New York, NY, USA, pp. 1402–1409.

32. Bassett, D. S., Owens, E. T., Porter, M. A., Manning, M. L. & Daniels, K. E. (2015) Extraction of force-chain network architecture in granular materials using community detection. *Soft Matter*, **11**, 2731–2744.

33. Barrat, A., Barthelemy, M. & Vespignani, A. (2008) *Dynamical Processes on Complex Networks*. Cambridge University Press.

34. Balcan, D., Hu, H., Gonçalves, B., Bajardi, P., Poletto, C., Ramasco, J. J., Paolotti, D., Perra, N., Tizzoni, M., Van den Broeck, W., Colizza, V. & Vespignani, A. (2009) Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Med.*, **7**, 45.

35. Colizza, V., Pastor-Satorras, R. & Vespignani, A. (2007) Reaction-diffusion processes and metapopulation models in heterogeneous-networks. *Nat. Phys.*, **3**, 276–282.

36. Xia, Y., Bjornstad, O. N. & Grenfell, B. T. (2004) Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Nat.*, **164**, 267–281.

37. Holme, P. (2015) Modern temporal network theory: a colloquium. *Eur. Phys. J. B*, **88**, 234.

38. DeDomenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gómez, S. & Arenas, A. (2013) Mathematical formulation of multilayer networks. *Phys. Rev. X*, **3**, 041022.

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

40. Newman, M. E. J. (2004) Analysis of weighted networks. *Phys. Rev. E*, **70**, 056131.

41. Reichardt, J. & Bornholdt, S. (2006) Statistical mechanics of community detection. *Phys. Rev. E*, **74**, 016110.

42. Onnela, J.-P., Fenn, D. J., Reid, S., Porter, M. A., Mucha, P. J., Fricker, M. D. & Jones, N. S. (2012) Taxonomies of networks from community structure. *Phys. Rev. E*, **86**, 036104.

43. Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z. & Wagner, D. (2008) On modularity clustering. *IEEE Trans. Knowl. Data Eng.*, **20**, 172–188.

44. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008) Fast unfolding of communities in large networks. *J. Stat. Mech.*, **2008**, P10008.

45. Jutla, I. S., Mucha, P. J. & Jeub, L. (2014) GenLouvain: a generalized Louvain method for community detection implemented in Matlab, version 2.0, 2011–2014. http://netwiki.amath.unc.edu/GenLouvain.

46. Lancichinetti, A. & Fortunato, S. (2012) Consensus clustering in complex networks. *Sci. Rep.*, **2**, 794.

47. Newman, M. E. J. (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, **103**, 8577–8582.

48. Zipf, G. K. (1946) The P_1P_2/D hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.*, **11**, 677–686.

49. Stewart, J. Q. (1947) Empirical mathematical rules concerning the distribution and equilibrium of population. *Geogr. Rev.*, **37**, 461–485.

50. Stewart, J. Q. & Warntz, W. (1958) Physics of population distribution. *J. Reg. Sci.*, **1**, 99–121.

51. Wilson, A. G. (1958) A statistical theory of spatial distribution models. *Transport. Res.*, **1**, 253–269.

52. Lenormand, M., Bassolas, A. & Ramasco, J. J. (2015) Systematic comparison of trip distribution laws and models. arXiv:1506.04889.

53. Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J. & Vespignani, A. (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA*, **106**, 21484–21489.

54. Lee, S. H., Ffrancon, R., Abrams, D. M., Kim, B. J. & Porter, M. A. (2014) Matchmaker, matchmaker, make me a match: migration of populations via marriages in the past. *Phys. Rev. X*, **4**, 041009.

55. Disdier, A.-C. & Head, K. (2008) The puzzling persistence of the distance effect on bilateral trade. *Rev Econ. Stat.*, **90**, 37–48.

56. Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L. & Daszak, P. (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.

57. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. (2012) A universal model for mobility and migration patterns. *Nature*, **484**, 96–100.

58. Goh, S., Lee, K., Park, J. S. & Choi, M. Y. (2012) Modification of the gravity model and application to the metropolitan Seoul subway system. *Phys. Rev. E*, **86**, 026102.

59. Masucci, A. P., Serras, J., Johansson, A. & Batty, M. (2013) Gravity versus radiation models: on the importance of scale and heterogeneity in commuting flows. *Phys. Rev. E*, **88**, 022812.

60. Stoddard, S. T., Morrison, A. C., Vazquez-Prokopec, G. M., Paz Soldan, V., Kochel, T. J., Kitron, U., Elder, J. P. & Scott, T. W. (2009) The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl. Trop. Dis.*, **3**, e481.

61. Simini, F., Maritan, A. & Néda, Z. (2013) Human mobility in a continuum approach. *PLoS One*, **8**, e60069.

62. Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D. & Woolrich, M. W. (2011) Network modelling methods for FMRI. *NeuroImage*, **54**, 875–891.

63. Strehl, A., Ghosh, J. & Cardie, C. (2002) Cluster ensembles—A knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, **3**, 583–617.

64. Traud, A. L., Kelsic, E. D., Mucha, P. J. & Porter, M. A. (2011) Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.*, **53**, 526–543.

65. Kraskov, A., Stägbauer, H., Andrzejak, R. G. & Grassberger, P. (2005) Hierarchical clustering using mutual information. *Europhys. Lett.*, **70**, 278–284.

66. Radicchi, F. & Arenas, A. (2013) Abrupt transition in the structural formation of interconnected networks. *Nat. Phys.*, **9**, 717–720.

67. Chowell, G., Torre, C. A., Munayco-Escate, C., Suárez-Ognio, L., López-Cruz, R., Hyman, J. M. & Castillo-Chávez, C. (2008) Spatial and temporal dynamics of dengue fever in Peru: 1994–2006. *Epidemiol. Infect.*, **136**, 1667–1677.

68. Guzman, M. G., Halstead, S. B., Artsob, H., Buchy, P., Farrar, J., Gubler, D. J., Hunsperger, E., Kroeger, A., Margolis, H. S., Martínez, E., Nathan, M. B., Pelegrino, J., Simmons, C., Yoksan, S. & Peeling, R. W. (2010) Dengue: a continuing global threat. *Nat. Rev. Microbiol.*, **8**, S7–16.

69. Depradine, C. & Lovell, E. (2004) Climatological variables and the incidence of Dengue fever in Barbados. *Int. J. Environ. Health Res.*, **6**, 429–441.
70. Instituto Nacional de Estadística e Informática (INEI). (2011–2014) http://www.inei.gob.pe/.
71. Chowell, G., Cazelles, B., Broutin, H. & Munayco, C. V. (2011) The influence of geographic and climate factors on the timing of dengue epidemics in Perú, 1994–2008. *BMC Infect. Dis.*, **11**, 164.
72. Kochel, T. J., Watts, D. M., Halstead, S. B., Hayes, C. G., Espinoza, A., Felices, V., Caceda, R., Bautista, C. T., Montoya, Y., Douglas, S. & Russell, K. L. (2002) Effect of dengue-1 antibodies on American dengue-2 viral infection and dengue haemorrhagic fever. *Lancet*, **360**, 310–312.
73. Montoya, Y., Holechek, S., Caceres, O., Palacios, A., Burans, J., Guevara, C., Quintana, F., Herrera, V., Pozo, E. & Anaya, E. (2003) Circulation of dengue viruses in north-western Peru, 2000–2001. *Dengue Bull.*, **27**, 52–62.
74. Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M. & Grafton, S. T. (2011) Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. USA*, **108**, 7641–7646.
75. Mucha, P. J. & Porter, M. A. (2010) Communities in multislice voting networks. *Chaos*, **20**, 041108.
76. Macon, K. T., Mucha, P. J. & Porter, M. A. (2012) Community structure in the United Nations General Assembly. *Physica A*, **391**, 343–361.
77. Sengupta, A. M. & Mitra, P. P. (1999) Distributions of singular values for some random matrices. *Phys. Rev. E*, **60**, 3389–3392.
78. Fenn, D. J., Porter, M. A., Mucha, P. J., McDonald, M., Williams, S., Johnson, N. F. & Jones, N. S. (2012) Dynamical clustering of exchange rates. *Quant. Finance*, **12**, 1493–1520.
79. Tatem, A. J., Rogers, D. J. & Hay, S. I. (2006) Global transport networks and infectious disease spread. *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications* (A. G. S. I. Hay and D. J. Rogers, eds.). Advances in Parasitology, vol. 62. Academic Press, pp. 293–343.
80. Truscott, J. & Ferguson, N. M. (2012) Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Comput. Biol.*, **8**, e1002699.
81. Anselin, L. (1995) Local indicators of spatial association-LISA. *Geogr. Anal.*, **27**, 93–115.
82. Kulldorff, M. (1997) A spatial scan statistic. *Commun. Stat. A–Theor*, **26**, 1481–1496.
83. GADM. (2015) Global Administrative Areas. http://www.gadm.org.
84. Thiemann, C., Theis, F., Grady, D., Brune, R. & Brockmann, D. (2010) The structure of borders in a small world. *PLoS One*, **5**, e15422.
85. Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R. & Strogatz, S. H. (2010) Redrawing the map of Great Britain from a network of human interactions. *PLoS One*, **5**, e14248.
86. Geonames.org (2011–2014) http://www.geonames.org/.
87. Kuno, G. (1997) Factors influencing the transmission of dengue viruses. *Dengue and Dengue Hemorrhagic Fever* (D. J. Gubler and G. Kuno, eds.). Wallingford, UK: CAB International, pp. 61–88.
88. Keating, J. (2001) An investigation into the cyclical incidence of dengue fever. *Soc. Sci. Med.*, **53**, 1587–1597.
89. Chowell, G. & Sanchez, F. (2006) Climate-based descriptive models of dengue fever: the 2002 epidemic in Colima, Mexico. *J. Environ. Health*, **68**, 40–44, 55.
90. Cummings, D. A. T., Irizarry, R. A., Huang, N. E., Endy, T. P., Nisalak, A., Ungchusak, K. & Burke, D. S. (2004) Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature*, **427**, 344–347.
91. Direcciòn General de Epidemiología (2011–2014) http://www.dge.gob.pe.
92. Fogarty International Center, National Institutes of Health (2011–2012) http://www.origem.info/misms/index.php.
93. Meilá, M. (2007) Comparing clusterings—An information based distance. *J. Multivariate Anal.*, **98**, 873–895.

Appendix A. Spatial benchmarks: quantifying the similarity of community structures using normalized variation of information and z-Rand scores

One can use many different diagnostics to measure the amount of similarity between planted partitions and algorithmically detected partitions. In our discussion of spatial benchmarks in Section 3, we used NMI to evaluate similarity. NMI is one of many normalized versions of mutual information (MI) [93]. Both MI and NMI are based on the concept of information entropy, which is a measure of uncertainty. MI measures the amount of information that one can predict about one random variable (which in the present paper is a partition of a network into communities) based on another one. For a partition $X = \{X_1, X_2, \ldots, X_K\}$ with K communities and a partition $Y = \{Y_1, Y_2, \ldots, Y_L\}$ with L communities, MI is defined as

$$I(X,Y) = \sum_{k=1}^{K} \sum_{l=1}^{L} P(k,l) \log_2 \left[\frac{P(k,l)}{P(k)P(l)} \right], \tag{A.1}$$

where $P(k)$ and $P(l)$ are the marginal probabilities of observing communities k and l in partitions X and Y, respectively, and $P(k,l)$ is the joint probability of observing communities k and l simultaneously in partitions X and Y. MI takes values between 0 and $\min\{H(X), H(Y)\}$, where $H(X) = -\sum_{k=1}^{K} P(k) \log_2 P(k)$ is the entropy of X.

NMI is given by the formula [63]

$$\mathrm{NMI}(X,Y) = \frac{I(X,Y)}{\sqrt{H(X)H(Y)}} \in [0,1]. \tag{A.2}$$

The normalization to lie within the range [0, 1] facilitates interpretation and comparisons.

In this appendix, we repeat our examination of the static uniform-population distance spatial benchmarks using normalized variation of information (NVI) [93] and z-Rand scores instead of NMI. In contrast to NMI, variation of information (VI) and NVI are metrics in the mathematical sense, and this is clearly desirable in some situations. See the discussions in Refs. [64,65,93]. VI and NVI are related to mutual information. For the partitions X and Y, VI is defined as

$$\mathrm{VI}(X,Y) = H(X) + H(Y) - 2I(X,Y) = H(X,Y) - I(X,Y), \tag{A.3}$$

where $H(X,Y) = H(X) + H(Y) - I(X,Y)$ is the joint entropy. VI is equal to 0 if the partitions X and Y are identical, and $VI(X,Y) \leqslant \log_2 N$, where N is the number of nodes in the whole network. We normalize VI to yield NVI [65]:

$$\mathrm{NVI}(X,Y) = 1 - \frac{I(X,Y)}{H(X,Y)} \in [0,1]. \tag{A.4}$$

We compare NMI and NVI with the z-score of the Rand coefficient [64], which is the diagnostic that we used for comparing algorithmic partitions of the disease-correlation networks to manual partitions. The z-Rand scores are good at detecting similarities in coarse structure [10,64], but (by design) they are not intended for discerning minor differences, such as one node changing community assignment. The Rand coefficient is

$$R = (w_{11} + w_{00})/M, \tag{A.5}$$

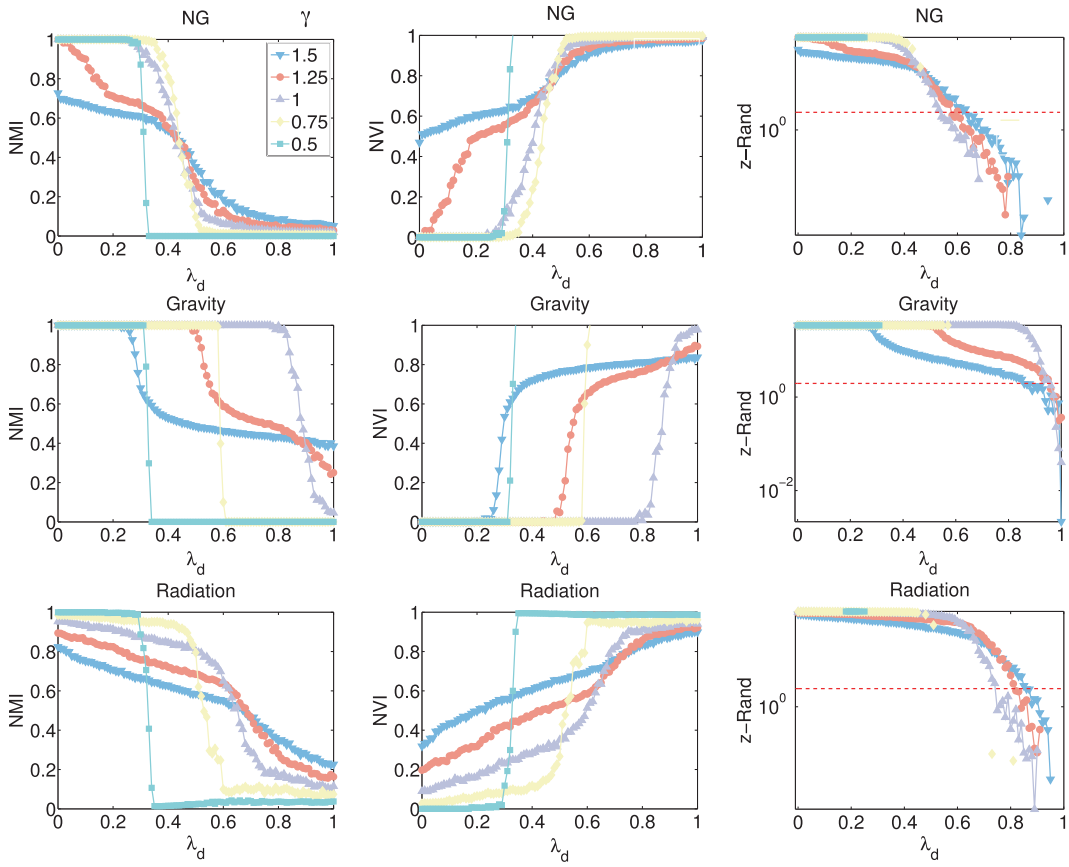

FIG. A1. (Left) Normalized mutual information (NMI), (centre) normalized variation of information (NVI) and (right) z-Rand scores (on a base-10 logarithmic scale) between algorithmically detected partitions, which we obtain by maximizing modularity, and planted partitions in the static uniform-population distance spatial benchmarks with $N = 50$ cities, an $l \times l$ grid with $l = 10$, and a density parameter of $\mu = 50$. We examine the partitions for different values of the resolution parameter γ as a function of inter-community connectivity λ_d using the (top) NG null model, (middle) gravity null model and (bottom) radiation null model. For the z-Rand scores, we show a significance cutoff of $z_R = 1.96$ (i.e., the 95% significance cutoff) for guidance. We do not plot values of 'Not a number' for partitions in which all nodes are assigned to one community.

where w_{11} is the number of node pairs that are assigned to the same community in both partitions, w_{00} is the number of node pairs that are assigned to different communities in both partitions and M is the total number of node pairs. Calculating a z-score of the Rand coefficient then yields a standardized measure of similarity [64].

As one should expect (given the relationship between the definitions) and one can see in Fig. A1, NMI and NVI give very similar results. The NMI and NVI values both change rather sharply at values of λ_d once the algorithmic partitions start to differ significantly from the planted partitions. In contrast, the z-Rand scores continue to construe the algorithmic partitions as similar to the planted partitions for larger values of inter-community connectivity λ_d until the deterioration of algorithmic partitions into singletons.

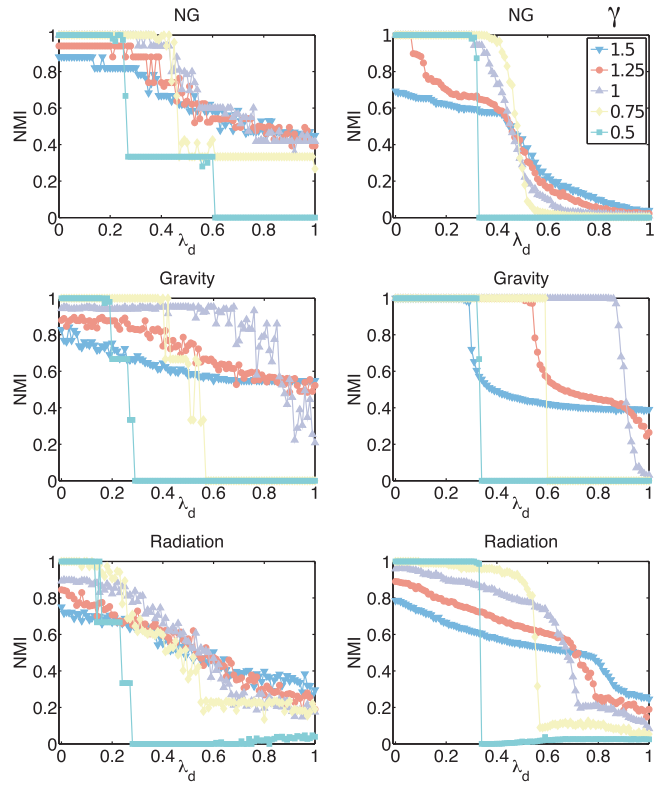

FIG. B1. Uniform-population static benchmarks: NMI scores between algorithmically detected partitions, which we obtain by maximizing modularity, and planted partitions in static uniform-population distance benchmarks with $l = 10$, a density parameter of $\mu = 100$, and uniform populations of 100 for different numbers of cities. (Because $l = 10$, the space in which we place the cities is the same in the two cases.) The numbers of cities are (left) $N = 10$ and (right) $N = 90$. We use the (top) NG, (middle) gravity and (bottom) radiation null models. See Fig. 3 in the main text for plots with $N = 50$.

Appendix B. Spatial benchmarks: varying the number N of cities

We now vary the number N of cities in benchmarks on an $l \times l$ grid with $l = 10$, a density parameter of $\mu = 100$ and a uniform population of 100 people in each city. In Fig. B1, we plot the NMI of algorithmic partitions versus planted partitions for several values of the resolution parameter γ for modularity maximization using the NG null model and both spatial null models. In combination with Fig. 3 in the main text, which has $N = 50$ cities, we observe no qualitative changes in NMI aside from an expected increase in variability when N is small.

Appendix C. Varying the edge-density parameter μ

We now present results of varying the edge-density parameter μ in static benchmarks. Edge density has a strong effect on the ability of modularity maximization to detect communities. For $\mu \lessapprox 5$, we obtain smaller NMI scores than we do for larger μ values. (See Figs. C1 and C2.) Thus, our calculations

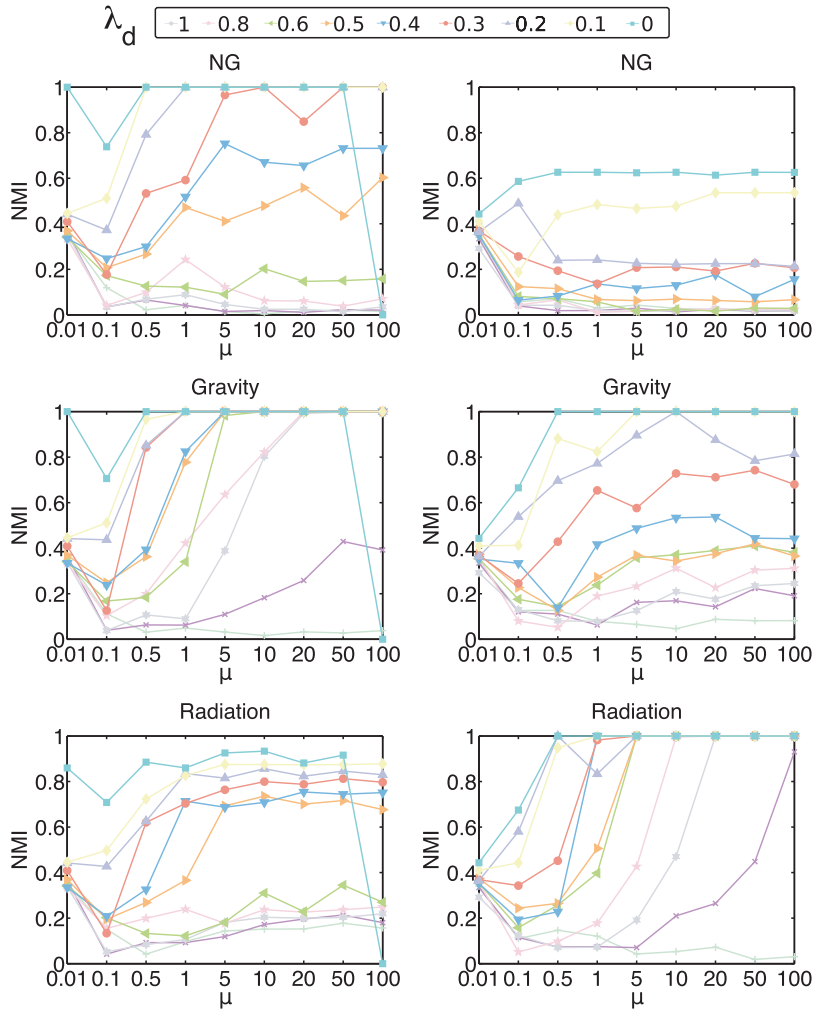

FIG. C1. NMI between algorithmically detected partitions, which we obtain by maximizing modularity with $\gamma = 1$, and planted partitions for static uniform-population spatial benchmarks with $N = 50$, a lattice-size parameter of $l = 10$, uniform city populations of 100 and several values of inter-community connectivity λ_d. We plot the NMI scores as a function of the edge-density parameter μ for (left) the distance benchmark and (right) the flux benchmark.

perform better on the benchmark networks for larger values of μ. To follow the choice that was used for the benchmark networks in Ref. [21], we used $\mu = 100$ in the main text.

Appendix D. 'Distance-and-population' benchmark

In Fig. 3 in the main text, we observed that the gravity null model for modularity maximization has the best performance among the null models for the uniform-population distance benchmark, but the NG null model performs better than spatial null models for the random-population distance benchmark because the edge placement in that benchmark does not include population information. We now

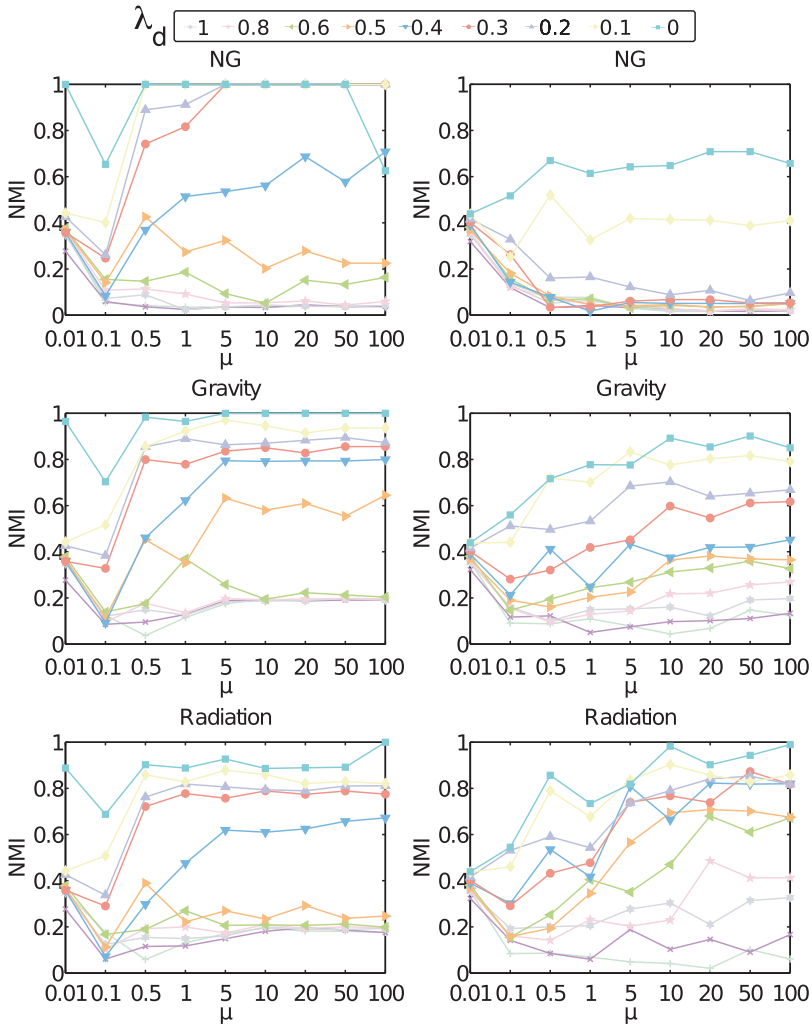

FIG. C2. NMI between algorithmically detected partitions, which we obtain by maximizing modularity with $\gamma = 1$, and planted partitions for static random-population spatial benchmarks with $N = 50$, a lattice-size parameter of $l = 10$, city populations n selected uniformly at random from $\{1, \ldots, 100\}$ and several values of inter-community connectivity λ_d. We plot the NMI scores as a function of the edge-density parameter μ for (left) the distance benchmark and (right) the flux benchmark.

study the effects of incorporating population into edge probabilities in what we call a 'distance-and-population' benchmark.

We construct our new type of benchmark network in the same manner as the distance benchmark in Section 3, but we now incorporate population into the edge-placement probability by taking $p_{ij}^{\text{distpop}} = n_i n_j \lambda(c_i, c_j)/(Z_3 d_{ij})$, where we recall that n_i and n_j are the populations of nodes i and j, the quantity d_{ij} is the distance between nodes i and j, the set c_i is the community that contains node i, the set c_j is the community that contains node j, and the function $\lambda(c_i, c_j) = 1$ if nodes i and j are in the same community and $\lambda(c_i, c_j) = \lambda_d$ otherwise. We also recall that the inter-community connectivity λ_d

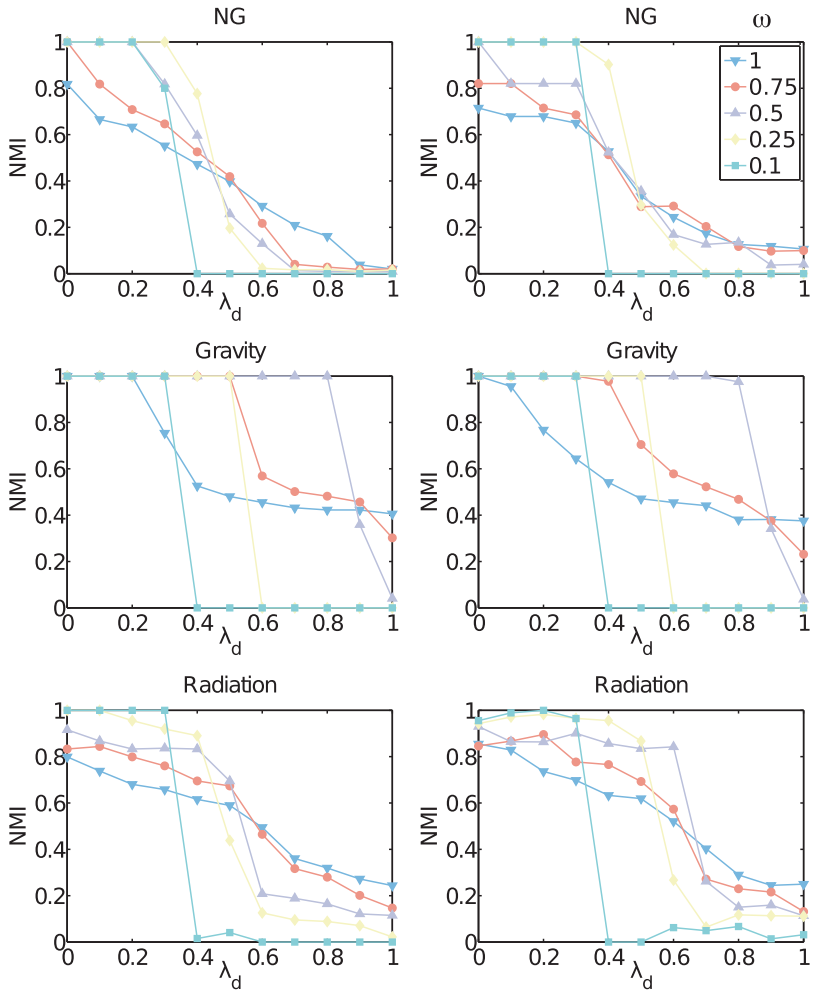

FIG. D1. NMI between algorithmically detected community structure, which we obtain by maximizing modularity, and planted community structure in 'distance-and-population' static spatial benchmarks with (left) uniform populations and (right) random populations. As in our prior examples, the population of each node is 100 in the uniform case, and it is drawn uniformly at random from the set $\{1, \ldots, 100\}$ in the random case. We use $N = 50$, $l = 10$, $m = 10$, $\mu = 100$ and $\gamma = 1$ for various values of ω (coloured curves) as a function of λ_d. We detect communities by maximizing modularity using the (top) NG, (middle) gravity and (bottom) radiation null models.

controls the amount of mixing between communities. When $\lambda_d = 0$, only nodes in the same community are adjacent to each other; when $\lambda_d = 1$, there are no distinct communities. The normalization constant Z_3 ensures that $\sum_{i>j} p_{ij}^{\text{distpop}} = 1$.

As expected, the gravity null model outperforms the other null models for the distance-and-population benchmark, which we recall was also the case for the uniform-population distance benchmark. (Compare Fig. D1 with Fig. 3 in the main text.) The radiation null model has the second-best performance on this benchmark, and its performance is better than it is on the random-population distance benchmark. However, it does not do as well as it does on the random-population flux benchmark (see Fig. 3).

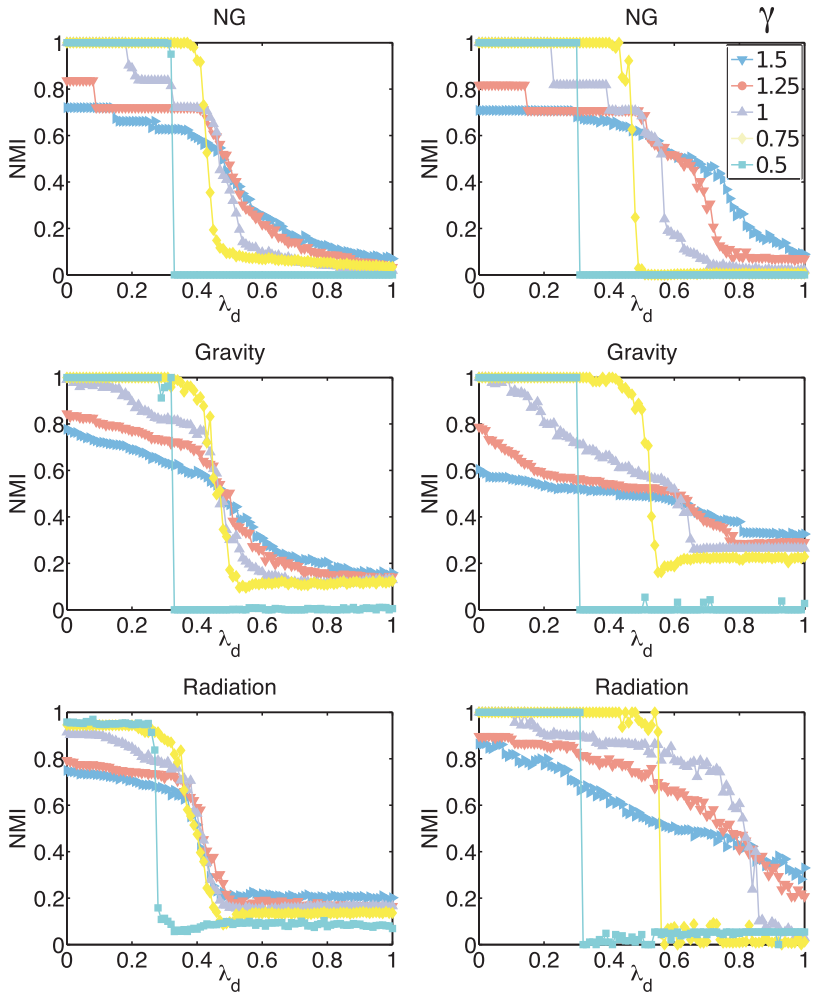

FIG. E1. NMI between algorithmically detected community structure, which we obtain by maximizing modularity, and planted community structure in multilayer random-population, temporally stable spatial benchmarks. We choose the population of each of the $N = 50$ cities uniformly at random from the set $\{1, \ldots, 100\}$. We consider various values of the resolution parameter γ, and the other parameter values are $l = 10$, $m = 10$, $\mu = 100$ and $\omega = 0.1$. We plot NMI as a function of λ_d for (left) the distance benchmark and (right) the flux benchmark using the (top) NG, (middle) gravity and (bottom) radiation null models.

Appendix E. Community detection in multilayer random-population, spatial benchmarks

We now study the influence of the parameters γ and ω on community structure in multilayer random-population, temporally stable benchmarks. We first compare the results with our findings from static benchmarks by varying γ and λ_d for fixed values of ω. We study the performance of modularity maximization with the NG, gravity and radiation null models on random-population benchmarks with $N = 50$ nodes, a lattice parameter of $l = 10$, and $m = 10$ layers using $\gamma \in \{0.5, 0.75, 1, 1.25, 1.5\}$ and $\omega \in \{0.1, 0.25, 0.5, 0.75, 1\}$. We only show plots for $\omega = 0.1$, as the values of ω do not noticeably influence the results.

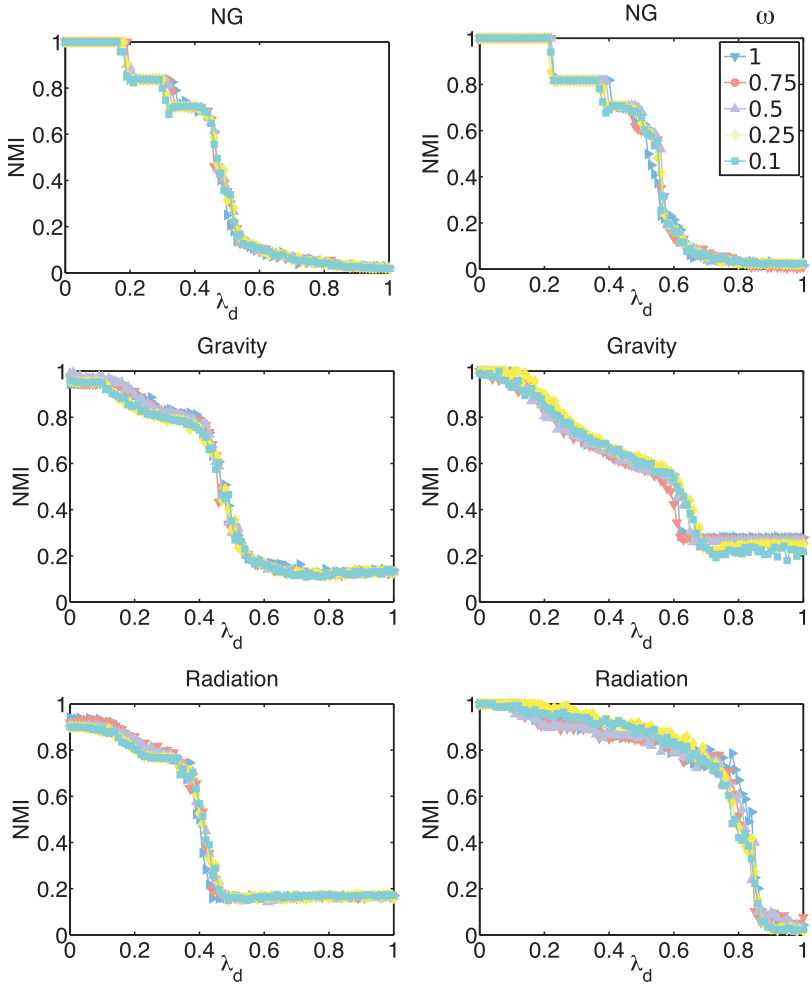

FIG. E2. NMI between algorithmically detected community structure, which we obtain by maximizing modularity, and planted community structure in multilayer random-population, temporally stable spatial benchmarks. We choose the population of each of the $N = 50$ cities uniformly at random from the set $\{1, \ldots, 100\}$. We consider various values of the parameter ω, and the other parameter values are $l = 10$, $m = 10$, $\mu = 100$ and $\gamma = 1$. We plot NMI as a function of λ_d for (left) the distance benchmark and (right) the flux benchmark using the (top) NG, (middle) gravity and (bottom) radiation null models.

We obtain results that are similar to the results that we showed for the corresponding static benchmarks in Fig. 3. Once again, we find that the choice of γ has a large influence on the quality of algorithmic partitions, and (as with our findings for static benchmarks) that $\gamma = 1$ seems to yield the best performance (i.e., the largest NMI scores) for low values of λ_d, whereas larger values of γ perform better for larger λ_d (see Fig. E1). In these experiments, the effect of varying γ is most pronounced when using the radiation null model on flux benchmarks.

We now examine the NMI of algorithmic versus planted partitions in multilayer temporally stable benchmarks for fixed $\gamma = 1$ while varying ω and λ_d. As we show in Fig. E2, we find that the value of ω usually has very little effect on our ability to detect planted communities via modularity maximization.

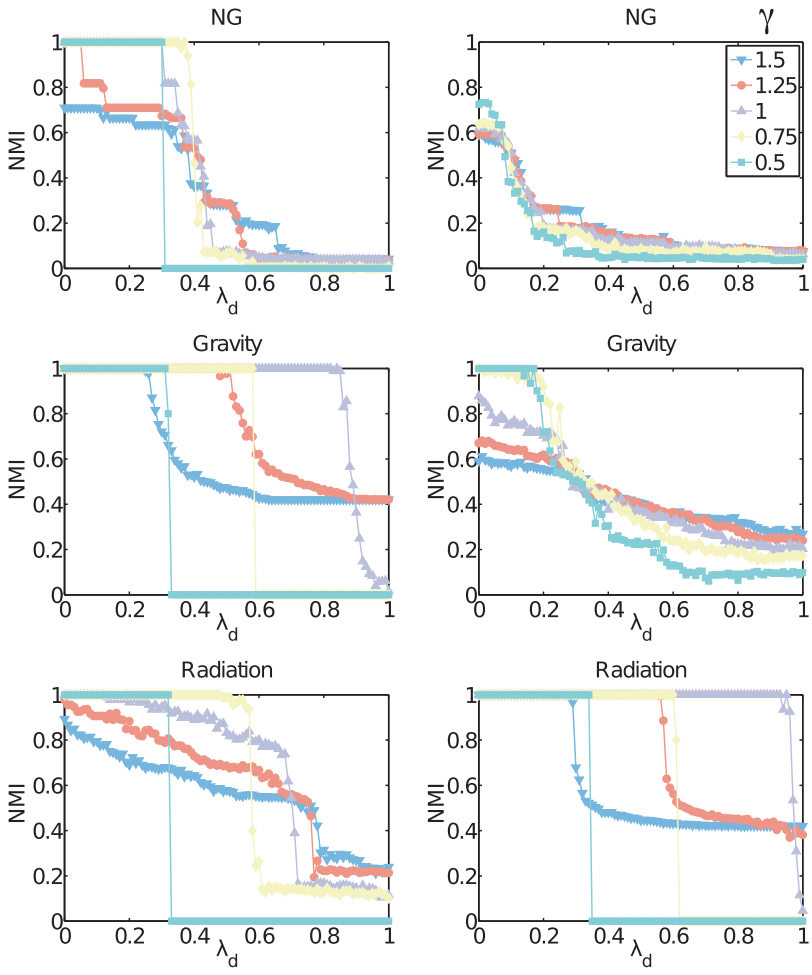

FIG. F1. NMI between algorithmically detected province-level community structure, which we obtain by maximizing modularity, for uniform-population ($n_i = 100$ for all i), temporally stable multilayer spatial benchmarks with $m = 10$ layers. Each layer has a single-layer planted partition with $N = 50$ cities, a lattice-size parameter of $l = 10$ and a density parameter of $\mu = 100$. We use $\omega = 0.1$ and consider various values of the resolution parameter γ, and we plot NMI as a function of the inter-community connectivity λ_d for (left) the distance benchmark and (right) the flux benchmark.

We observed the same scenario for the multilayer uniform-population, temporally stable benchmarks (see Fig. 5 in the main text). However, ω is important for the random-population, temporally evolving multilayer benchmarks, which is also what we observed in the main text for uniform-population benchmark networks. See Fig. 6 in the main text for an illustration in the uniform-population case. We do not show an analogous figure for the temporally evolving case.

Appendix F. Province-level communities for multilayer benchmarks

In Fig. F1, we present our results for province-level community structure on multilayer uniform-population, temporally stable benchmarks. As one can see by comparing these results to those in Fig. 4

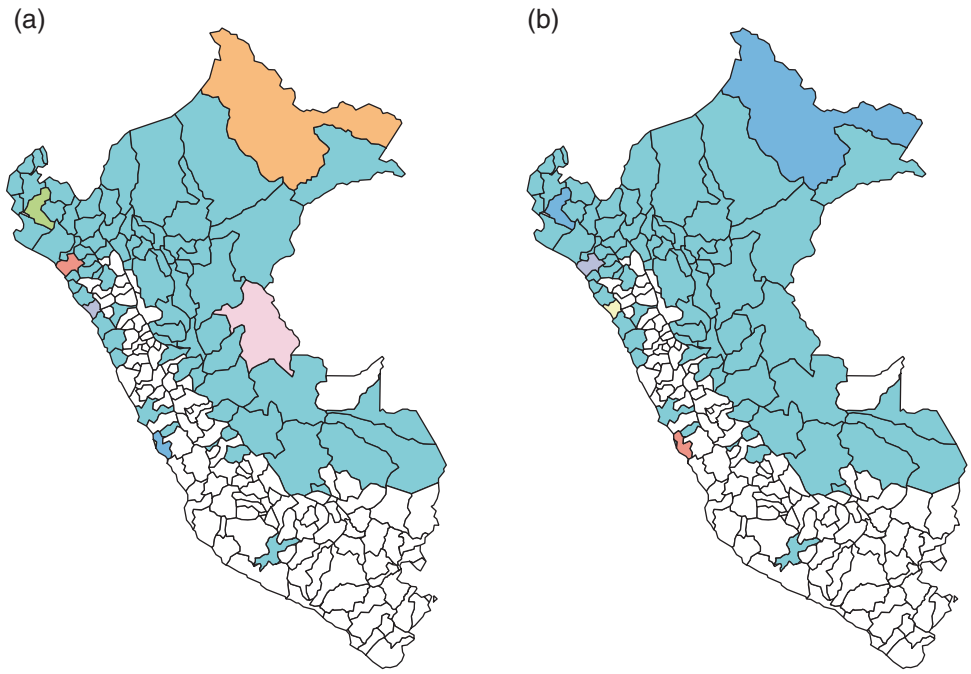

FIG. G1. Algorithmically detected community structure, which we obtain via modularity maximization (with a resolution-parameter value of $\gamma = 1$), in static dengue fever correlation networks that we construct using the entire set of time series (i.e., we use $\tau = 1$ and $\Delta = 779$). We show partitions from using (a) the gravity null model and (b) the radiation null model. We colour provinces on a map of Peru according to their community assignment. White provinces are ones in which our data does not include any reported cases of dengue fever.

in the main text, we obtain similar NMI scores for the performance of community detection for province-level communities as we did for ordinary community detection in multilayer networks.

Appendix G. Community structure in aggregated dengue fever data

In Fig. G1, we show additional results of community detection on fully aggregated networks (i.e., we use $\tau = 1$ and $\Delta = 779$) from the dengue fever time series. In Section 4.3.3 of the main text [see Fig. 13(a)], we showed the results of modularity maximization using the NG null model. We now show our results for this type of experiment by maximizing modularity using the gravity and radiation null models. The gravity and radiation null models yield one large community and a few small communities. Because of the aggregation, we have lost the rich set of information that we were able to study using multilayer community detection.