Spatial Strength Centrality and the Effect of Spatial Embeddings on Network Architecture

Andrew Liu

Department, of Mathematics, University of California, Los Angeles, 90095, USA

Mason A. Porter

Department, of Mathematics, University of California, Los Angeles, 90095, USA

(Dated: August 25, 2019)

For many networks, it is useful to think of their nodes as being embedded in a latent space, and such embeddings can affect the probability of whether nodes are adjacent to each other. In this paper, we extend existing models of synthetic networks to spatial network models by incorporating deterrence functions. We start by extending an existing geographical fitness model by employing normally-distributed fitnesses, and we then develop spatial versions of preferential attachment and configuration models. We define a notion of "spatial strength centrality" to help characterize how strongly a spatial embedding affects network structure, and we examine spatial strength centrality on a variety of real and synthetic networks.

map: various things to do:

1. **[MAP's note to self on some papers to cite:] I would like to cite a paper by Eleni Katifori, if possible (it seems that some of her stuff is relevant); I also want to cite one of Geoff Boeing's papers at some point [7/30/19: I'll now do this next time, as I am seeing sufficient imprecision and confusion in parts of the paper that those parts will need clarification, so I may as well add these final citations in the next round instead]; I also want to cite our paper Berthier et al (PNAS 2019), as we consider spatial ideas rather explicitly therein (including trying to separate geometry and topology); citing our about-to-go-live article Nauer et al. probably also makes sense**

2. **please make changes to text in a new color so that I can just go through the changed parts of the paper when I see this again**

I. INTRODUCTION

Many networks have important spatial features, and it is often useful to think of the nodes of such spatial networks as being embedded in either a physical or a latent space [1, 2]. Such a latent space can be literal in nature, like the physical distance between different parts of a city, or they can be more abstract. In social networks, for example, one can construe the nodes that represent individuals as having a distance between them that represents physical distance; a distance between them that arises from demographic characteristics, interests, behaviors, or other features; or a combination of such distances [3]. In food webs, nodes that represent individuals can be embedded in a latent space that represents various features and affects their likelihood of interacting with each other. For instance, in a niche model, species are placed on a line, with a position that represents their mass and directed edges between nodes to represent who eats whom [4].

Physical networks, such as transportation networks and other networks within and between cities, possess a natural spatial embedding based on their physical location in the world. Whether networks are embedded in an explicit or latent space (or both), the distances between nodes can strongly influence whether they are adjacent to each other in a network [5–7]. Intuitively, nodes that are farther apart from each other have fewer opportunities to interact (or their interaction has a greater associated cost), so we expect that there is a concomitant lower probability to observe edges between them. For example, people in a social network may be unlikely to interact if they share little in common [8], and animals in a food web may not hunt others that are too large or small in comparison to them [4].

It is sensible that spatial characteristics should influence network structure, but it is much more difficult to make these ideas precise [1]. For example, how exactly does a spatial embedding influence the architecture — both topology and edge weights — of a given network? For simplicity, most studies of networks have ignored spatial embeddings, but new models that incorporate spatial considerations are being developed with increasing frequency [1].

Network models that incorporate space can improve measures of importance (i.e., centrality) of networks that are constructed from empirical data [5, 6]. Additionally, spatially relevant features have been examined in brain networks [9, 10], fungal networks [11], road networks [5, 12, 13], air-traffic networks [6], gas-pipeline networks [14], and other applications. Incorporating spatial features can improve the modeling of empirical data [1], and it is therefore important to further extend ideas from network analysis into the realm of spatial networks.

Examples of spatial models include spatially-embedded

random networks [15, 16] and networks that grow in time (e.g., through preferential attachment) [5, 17, 18]. Such models can provide reference models with which to compare empirical networks; and ideas from spatial networks have also led to spatial null models for community detection [19, 20]. Some spatial network models produce networks with multiple properties that are reminiscent of empirical networks (such as simultaneously having large values of clustering coefficients and heavy-tailed degree distributions [21, 22]). Spatial network models have also been helpful for studying models of biological growth, such as osteocyte-network formation [17].

map: I wonder if the sentence above would be a good place to cite one of Eleni Katefori's papers

map: in terms of urban structures (cities), I think we should cite Geoff Boeing at some point too

In the present paper, we explore a general approach for extending non-spatial network models to spatial versions by introducing a deterrence function [20]. This deterrence function $h(r)$, where r represents the distance in either latent or physical space, modifies the probability that nodes are adjacent to each other as a function of the distance between them. We start by examining a modification of a geographical fitness (GF) model [23] that is a latent-space ("hidden variable") model. Other works examining this GF model [16, 22] have used exponential and power-law fitness functions, but we use one that is normally distributed. We then explore a spatial extension for the Barbábasi–Albert (BA) model, where this deterrence function modifies attachment probability for new edges (previous papers examine this [18, 24, 25], but we incorporate the deterrence function slightly differently). Finally we apply this deterrence function for a configuration model [26] to create a spatial configuration model. We compute some characteristics on these new spatial network models to understand how this deterrence function affects network topology. We also define a new spatial notion of centrality, which we call "spatial strength centrality"" that helps us measure how strongly a spatial embedding affects the topological structure of a network. We examine how spatial strength centrality behaves on our new synthetic models, and we compute it for a variety of empirical spatial networks.

Our paper proceeds as follows. In Section II, we discuss GF models and explore the properties of these networks when using normally distributed fitnesses. We modify a spatial preferential attachment (SPA) model in Section III and explore its properties, and develop a new spatial configuration model in Section IV. In Section V, we define our notion of spatial strength centrality and apply it to several empirical and synthetic networks, including the models that we explored in previous sections. We conclude in Section VI.

II. GEOGRAPHICAL FITNESS MODEL

A. Prior research

The (non-geographical) fitness network model is a class of networks that assigns fitness values to the n nodes of a network [23]. Such models are sometimes also called "threshold models" [16], although one needs to be careful not to confuse them with other similarly-named models [2]. One determines the intrinsic fitnesses of the nodes from a density function $f(w)$. In the original model [27], this intrinsic fitness characterizes the propensity of a node to gain edges. Such models have been used to generate small-world networks with power-law degree distributions [16].

One choice of fitness distribution is the exponential distribution

$$f(w) = \lambda e^{-\lambda w}, \quad w \geq 0. \tag{1}$$

It gives the probability that a node has an intrinsic fitness value of w, where the parameter $\lambda \geq 0$ determines the shape of the distribution. An edge exists between nodes v_i and v_j, with $i \neq j$, when

$$g(i, j) = w_i + w_j \geq \theta, \tag{2}$$

where θ is a "threshold parameter" that determines the fitness values that nodes need to be adjacent. With equation (2), nodes that have a higher fitness value w also have a larger degree. With either an exponential distribution or power-law distribution of fitness, one can show that the resulting degree distribution follows the power law $p(k) \sim k^{-2}$ as $k \to \infty$ [27, 28].

One can also formulate geographical (or, more generally, spatial) versions of a fitness model, as illustrated in [16, 27, 29]. In one extension, a network exists in a d-dimensional Euclidean space of finite size, such as in $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ for $d = 2$. In addition to assigning node fitnesses using the distribution $f(w)$, one now also assigns a location in space to each node v_i. For example, perhaps one determines each of the node's d coordinates uniformly at random in the space. One can then suppose that an edge exists between nodes v_i and v_j, with $i \neq j$, when

$$g(i, j, r) = (w_i + w_j)h(r) \geq \theta, \tag{3}$$

where r is the Euclidean distance between nodes v_i and v_j and the function $h(r)$ describes the influence of space on the connection probability of two nodes as a function of the distance between them. The function $h(r)$ is sometimes called a "deterrence function" [13], and it has been used in many studies of spatial networks [1]. A common choice for a deterrence function is the power-law form [16]

$$h(r) = r^{-\beta}, \tag{4}$$

for some value of a spatial decay parameter β. For certain values of β, it is possible to determine exact expressions

for degree distributions and global clustering coefficients for the resulting fitness-network model [16], but most investigations with this deterrence function have focused on numerical computations.

B. Normally-distributed fitnesses

In previous treatments of (geographical and non-geographical) fitness network models, exponential and power-law distributions are very common choices, including for the specific purpose of generating networks with power-law degree distributions [16, 22, 27, 29].

In the present study, we use a normal distribution for fitness, rather than an exponential or power-law distribution. We make this choice based on the intuition that entities (which are represented by nodes) have a variety of intrinsic factors that influence whether they interact with other nodes in a system, so this is a plausible choice in many situations [30]. We chose a standard normal distribution

$$W \sim N(0, 1). \tag{5}$$

We also modify the threshold function and write

$$g(i, j, r) = |w_i - w_j| h(r) \geq \theta, \tag{6}$$

such that nodes that differ widely in intrinsic fitness are more likely to interact with each other than those with similar intrinsic fitness. That is, we generate networks with disassortative mixing with respect to node fitness. Possibilities where this may be relevant may be hyperlinks from many low-fitness Web pages to high-fitness ones and predators in a food web hunting much weaker prey.

C. Network realizations and numerical computations

For our numerical computations, we construct networks with $n = 500$ nodes that are embedded in 2-dimensional (2D) space in a box $[0, 1] \times [0, 1]$ with periodic boundary conditions. The nodes have coordinates (x_1, x_2), where we assign x_1 and x_2 independently with uniform probability on the interval $[0, 1]$, and normally-distributed fitnesses with mean $\mu = 0$ and standard deviation $\sigma = 1$.

For the threshold function (4), we consider values $\beta \in \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$ for the decay parameter β, and we manually adjust θ such that the mean degree $\langle k \rangle = 20$. We obtain θ through a combination of trial-and-error and fitting through linear regression. We generate 30 instances of these networks for each value of β.

When $\beta = 0$, we recover a non-geographical threshold model, where the distances between nodes do not play a role in the probability of nodes to be adjacent. In this regime, the distribution looks visually like it may satisfy a power law, although we do not test this idea. For progressively larger values of β, distance plays a progressively stronger role, and nodes that are farther apart are less likely to be adjacent to each other.

We calculate a few well-known network characteristics [2] and find that they are affected by distance (as encoded in the value of β). The characteristics that we calculate are

1. **Mean local clustering coefficient.** The local clustering coefficient of a node with degree at least 2 is $c_i = \frac{2T(v_i)}{k_i(k_i - 1)}$, where $T(v_i)$ is the number of unique triangles (i.e., 3-cliques) to which node v_i belongs and k_i is the degree of node v_i. A node with degree 0 or 1 has a local clustering coefficient of 0. The mean local clustering coefficient of a network is the mean value of c_i over all nodes in the network (including nodes with degrees 0 and 1).

2. **Mean geodesic distance.** The mean geodesic distance is $L = \frac{\sum_{i \neq j} d(v_i, v_j)}{n(n-1)}$, where $d(v_i, v_j)$ is the shortest-path distance (in terms of number of edges) between nodes v_i and v_j.

3. **Mean edge length.** We take the length of an edge to be the Euclidean distance between its two incident nodes. The mean edge length of a network is the mean edge length over all edges in the network.

4. **Degree assortativity.** We calculate the degree assortativity of a network using the Pearson correlation coefficient $r = \frac{\sum_{i,j} (A_{i,j} - k_i k_j / 2m) k_i k_j}{\sum_{i,j} (k_i \delta_{i,j} - k_i k_j / 2m) k_i k_j}$ [2], the normalized covariance of degrees of the nodes of the network.

Larger values of β yield shorter mean edge lengths, larger mean geodesic lengths, and larger mean local clustering coefficients. In our plots in this section, we calculate the means of these characteristics in a network, and we then take the mean across 30 instantiations. We note that this result for clustering coefficient is intuitively sensible, as nodes in small neighborhoods of nodes are more likely to be adjacent to each other for larger values of β (i.e., as spatial effects become more prominent).

To facilitate exposition, we use the term "spatial effects" to describe network topology and network characteristics being influenced by a spatial embedding, such that greater spatial effects signify more influence of a spatial embedding on a network. With the deterrence function $h(r) = r^{-\beta}$, we expect spatial effects to increase as we increase β. The most obvious example of this phenomenon is with mean edge length, a network characteristic. With this model (and the subsequent ones in this paper), as we increase β, edges form preferentially between spatially close nodes, so mean edge length decreases as we increase β.

Near $\beta = 0$, we observe a lower degree assortativity because of the heterogeneous mixing of nodes with different
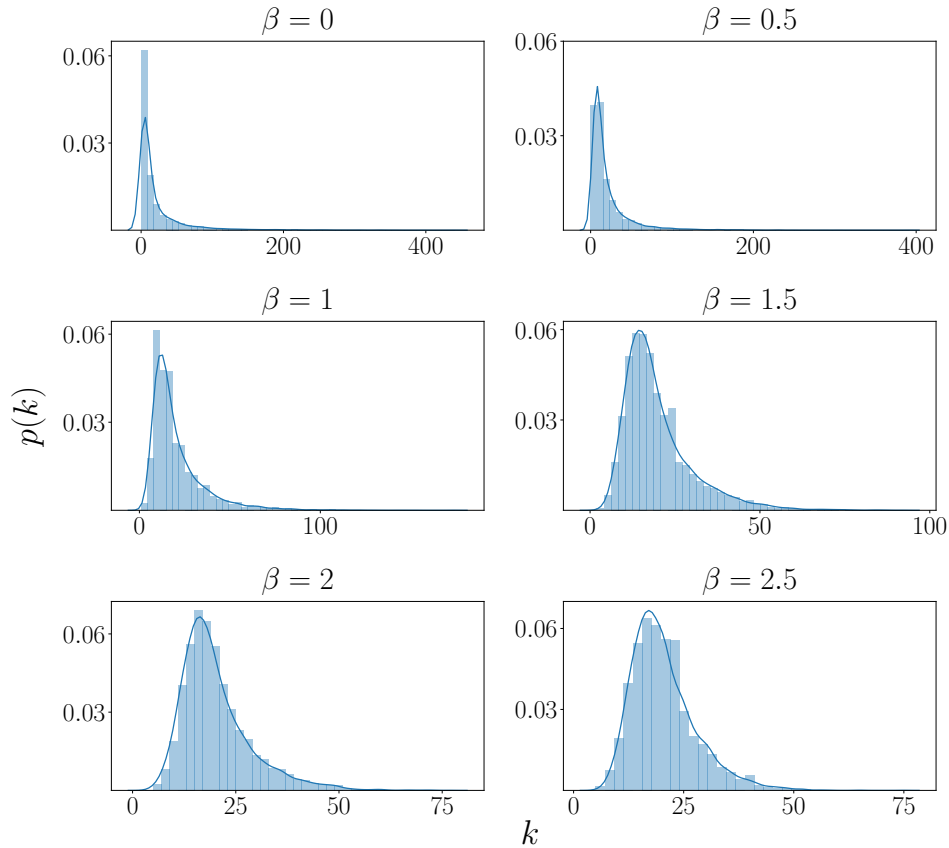
FIG. 1. Example degree distributions of geographical fitness networks with normally-distributed fitnesses and various values of the decay parameter β.

degrees. This is a consequence of how we formulated the model, as nodes with extreme fitness values (either positive or negative ones) yield larger values of $|w_i - w_j|$, so they are adjacent to all nodes with fitnesses that are sufficiently close to 0. These nodes, whose fitnesses are in the center of the fitness distribution, are not likely to be adjacent to each other.

By contrast, for larger values of the decay parameter β (e.g., for $\beta \geq 2$), the Euclidean distance between two nodes has more influence over the probability that they are adjacent. In this regime, it is no longer the case that nodes with extreme fitness are adjacent to all nodes in the network.

If we let $\beta \to \infty$, we expect the distance between nodes to become the sole factor that determines whether nodes are adjacent. In our numerical computations, we hold $\langle k \rangle$ roughly constant at 20, so we expect nodes to be adjacent to roughly 20 of their nearest neighbors. In this respect, as $\beta \to \infty$, we expect the network to resemble a random geometric graph (RGG) [15, 31] with an appropriate value of the connection radius r. The RGG considered in this paper is a graph model that is given two paramemters: the number of nodes n, and a connection radius r_c. Each node is added to a $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ with its location assigned uniformly at random in the space. Pairs of nodes are adjacent to each other if their Euclidean distance is less than or equal to r_c in the space. Thus adjacency in the network is determined completely by the position of its nodes.

For example, in an instantiation of this model with $\beta = 50$, an RGG (with periodic boundary conditions) with the same node positions and with $r \approx 11.3$ (thus having $\langle k \rangle \approx 20$) yields a network with 98% of the same edges. The same experiment with $\beta = 3$ yields 77% of the same edges and with $\beta = 1$ yields 52% of the same edges (percentages are precise to 2 digits). Random geometric graphs have positive degree assortativity [32], so as we increase β, we expect degree assortativity to increase to positive values that resemble those of this type of RGG.

D. Closeness centrality

Normalized closeness centrality gives an idea of how "close" a node is to other nodes in a network, as determined by how many edges are between it and other nodes in the network [2]. The version of closeness cen-
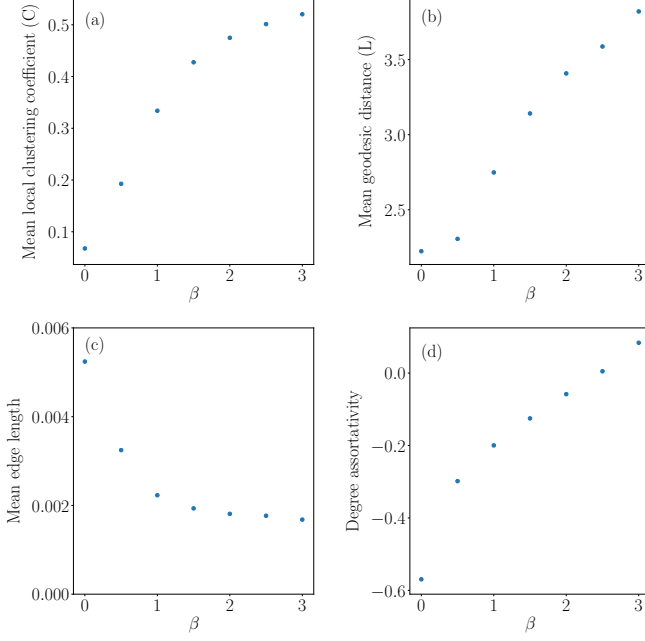
FIG. 2. Characteristics of networks (with $n = 500$ nodes) from a geographical fitness models with normally-distributed fitness for different values of β and 30 instantiations of the model for each value of β. We show the (a) mean local clustering coefficient, (b) mean geodesic distance, (c) mean edge length, and (d) degree assortativity. In all cases, we first take means with respect to individual networks and then over the 30 instantiations.

trality that we calculate is

$$ C(v_i) = \frac{n-1}{\sum_{i \neq j} l(v_i, v_j)} \,, \tag{7} $$

where $l(v_i, v_j)$ is the geodesic distance between nodes v_i and v_j.

For $\beta = 0$, the closeness centralities of the nodes arise only from their intrinsic fitness values, as the model does not incorporate any spatial effects. In the scatter plots of fitness versus closeness centrality, we observe two large branches of values (see Fig. 3) in all networks from our 30 instantiations. In these networks, the node whose fitness has the large absolute value (which we call the "extremum") is adjacent to all nodes in the network that have sufficiently different fitness; these nodes all belong to the other large branch, and the extremum acts as a shortcut to other nodes in the network.

In these scatter plots, we also observe a small central curve between the two branches. It lies below the lowest point of the two branches (such as in the bottom-right scatter plot of Fig. 3) or above the nearby parts of both branches. This curve is below the branches when no nodes in the network have fitnesses with large enough absolute values, such that the nodes in the curve are not adjacent to any other node. Otherwise, they are adjacent to nodes with very negative or very positive fitnesses.
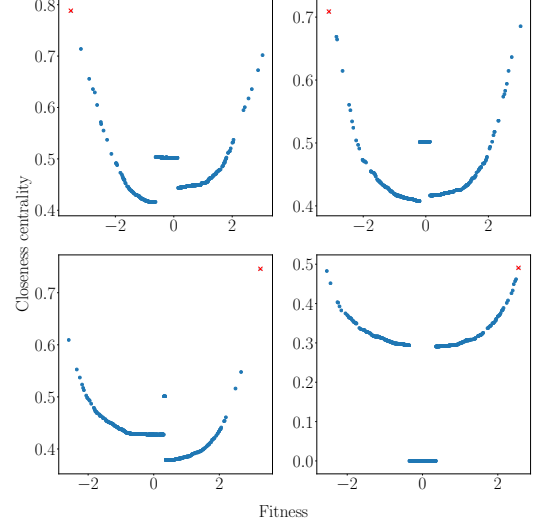

FIG. 3. Scatter plot of closeness centralities for nodes in 4 instantiations of the geographical fitness model with decay parameter $\beta = 0$. All instantiations for $\beta = 0$ have similar scatter plots. The depicted instantiations illustrate common patterns. For each network, we highlight the node with the largest (in absolute value) fitness. In most scatter plots, we observe two large branches (on the left and right) and a small horizontal (or predominantly horizontal) curve between these two large branches. We refer to one branch as "lower" than the other if the lowest point of that branch is lower than the lowest point of the other branch. We observe that the left branch is lower when there is an extremum on the left (see the top-left and top-right panels) and that the right branch is lower when there is an extremum on the right (see the bottom-left plot). If the node with fitness closest to zero has a fitness value of w_m and the extremum has a fitness value of w_M, then the small central curve is below the branches when $|w_M| < \theta - |w_m|$ (θ is approximately 2.9 for $\beta = 0$).

Extremum nodes still behave as shortcuts for $\beta = 0.5$, but the spatial constraints necessitate spatial proximity for nodes to be adjacent to other nodes, so the role of the extremum as a shortcut is less important, and some shortest paths go through other nodes. We show example scatter plots of closeness versus fitness for $\beta = 0.5$ in Fig. 4. In the top-left plot, for example, the presence of an extremum with very large fitness still yields some branches in the scatter plot. However, in the other examples in this figure (and in most scatter plots from the 30 instantiations), branches are less obvious. For $\beta = 1$ (and hence for larger values of β), branches are almost never apparent, as node fitnesses are less likely to determine whether two nodes are adjacent; instead, spatial effects become more prominent. We thereby see how spatial effects manifest in this geographical threshold model.
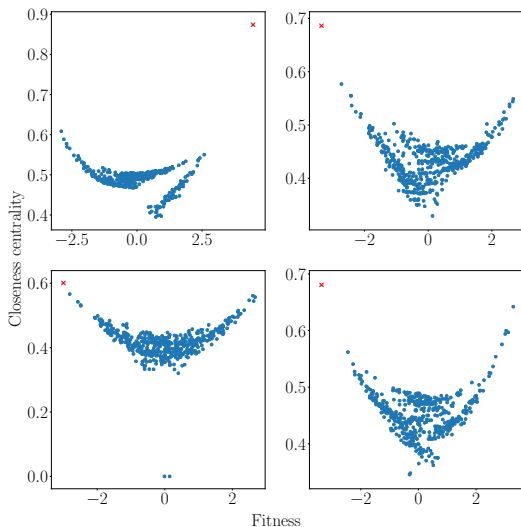
FIG. 4. Scatter plot of closeness centralities for nodes in 4 instantiations of the geographical fitness model with decay parameter $\beta = 0.5$. The depicted instantiations illustrate common patterns. For each network, we highlight the node with the largest (in absolute value) fitness. For this value of β, we do not always observe noticeable branches. (See the top-right, bottom-left, and bottom-right panels.) When there are noticeable branches in a scatter plot, they are usually accompanied by a node with a large absolute value of the fitness. (See the top-left panel.)
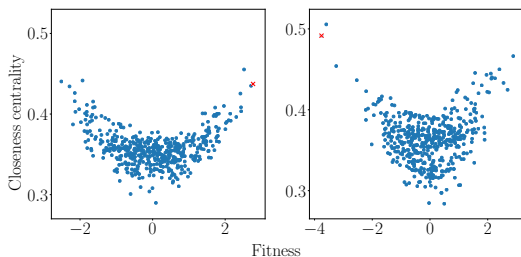

FIG. 5. Scatter plot of closeness centralities for nodes in 2 instantiations of the geographical fitness model with decay parameter $\beta = 1$. The depicted instantiations illustrate common patterns. For each network, we highlight the node with the largest (in absolute value) fitness. For this value of β, branches are almost never apparent in the scatter plots, even with extreme fitness value nodes (see the right plot), and the most extreme fitness value node is no longer always the one with the largest value of closeness centrality.

III. A SPATIAL PREFERENTIAL-ATTACHMENT MODEL

A. Prior research

We now examine characteristics of a spatial generalization of the BA preferential-attachment model [33]. Such a spatial preferential-attachment model (SPA) was first introduced in [18] and explored further in [24, 25, 34].

It starts with a seed network, with nodes located in $[0,1] \times [0,1] \subset \mathbb{R}^2$), at $t = 0$. During each subsequent time step, one adds a new node and assigns its location uniformly at random in the space. If a node is assigned the same location as an existing node, a different location is assigned uniformly at random. New nodes are connected to m existing nodes with probability

$$p(v_i, v_j) = \frac{k_j h(r_{i,j})}{\sum_l k_l h(r_{i,l})}, \qquad (8)$$

where $r_{i,j}$ is the Euclidean distance between nodes v_i and v_j and $h(r_{i,j})$ is the same as in (4) (i.e., $h(r_{i,j}) = r_{i,j}^{-\beta}$). In [18], this SPA was simulated on a 1D space. Based on numerical computations (taken as a mean over 10 realizations), it appeared that it might have a power-law degree distribution $P(k)$ for $\beta < 1$ and a stretched exponential degree distribution for $\beta \geq 1$. For $\beta < 1$, the authors noted deviations from the corresponding power-law for small values of k. Based on the same simulations, the mean geodesic distance of the network appeared to scale logarithmically with the time steps for a wide range of β (shown with up to $\beta = 5$). References [24, 25] examined slight modifications to this SPA model, using the same connection probability 8 and generation procedure, but embedding the network in 2D space. Reference [24] assigned one new edge per incoming node and explored the distribution of edge lengths and the expected degree of a node born at time t.

Other spatial generalizations of the BA model have also been studied. In [35, 36], for example, each node has a "sphere of influence" that's size is proportional to the node's in-degree. When new nodes are added, if their location is inside the sphere of influence for a given node, they have a probability p to create an edge to that node. As $t \to \infty$, the model exhibits power-law degree distribution and has a positive limit for the mean local clustering coefficient [36].

B. Description of our SPA model

In the present paper, in contrast to the SPA models in [18, 24, 25], which use a connection probability (8), we instead use a connection probability function of

$$p(v_i, v_j) = \frac{k_j}{\sum_l k_l} h(r_{i,j}). \qquad (9)$$

The intuition behind our choice is that the "popularity" of a node (given by its relative degree in a network) is independent of its distance from another node. This allows us to transform an existing non-spatial PA model (such as the BA model) into a spatial variant by multiplying the probability that an edge forms by a deterrence function $h(r)$. The probability that an incoming node v_t at time step t forms an edge with any other node in the network (after normalizing this probability such that it
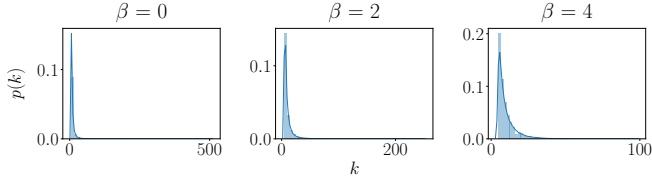
FIG. 6. Degree distributions for single instances of our SPA model for $T = 10000$ time steps. From left to right, we show examples for $\beta = 0$, $\beta = 2$, and $\beta = 4$.

sums to 1) is thus

$$p(v_t, v_j) = \frac{p(v_t, v_j)}{\sum\limits_{\substack{j < t \\ (j,t) \notin E}} p(v_t, v_j)}, \tag{10}$$

where E is the set of edges of the network. When implemented in an algorithm, the edges are assigned one at a time, and we do not consider adding edges to nodes v_t is already adjacent to, or v_t itself, thus disallowing self-edges and multi-edges.

We assign a location uniformly at random in $[0, 1] \times [0, 1]$ to each node as we add it to the network. We start our network with a seed that consists of a 10-clique (each node in this 10-clique is assigned a location uniformly at random). At each time step, we add a new node v_i with $m = 5$ stubs (i.e., ends of edges), and we connect each of these stubs to an existing node in the network with probability equal to (10). Consequently, nodes with larger degrees are more likely to accrue more connections and incoming nodes are more likely to connect to nearby nodes than to ones that are farther away. All edges are undirected and unweighted.

For computations, we consider decay parameters of $\beta \in \{0, 1, 2, 3, 4\}$, and we examine 10 instantiations of our model for each value of β. We simulate each instantiation of our model for T time steps, and we consider $T \in \{300, 1000, 3000, 10000\}$ (note that these will respectively have $n \in \{310, 1010, 3010, 10010\}$ nodes due to the 10 nodes in the seed network). Because each node adds 5 edges to a network, it follows that $\langle k \rangle \to 10$ as $T \to \infty$.

C. Computational results

As with the GF model, we calculate the mean local clustering coefficient, geodesic distance, edge length, and degree assortativity. (See Section II for definitions of these quantities.) In our SPA model, we observe the same general trends as we increase the spatial decay parameter β in the examined network characteristics as in the GF model. Specifically, the mean local clustering coefficient, geodesic distance, and degree assortativity increase as we increase β; and the mean edge length decreases as we increase β.

We examine in more detail how the mean geodesic distance (see Fig. 8) and mean local clustering coefficient
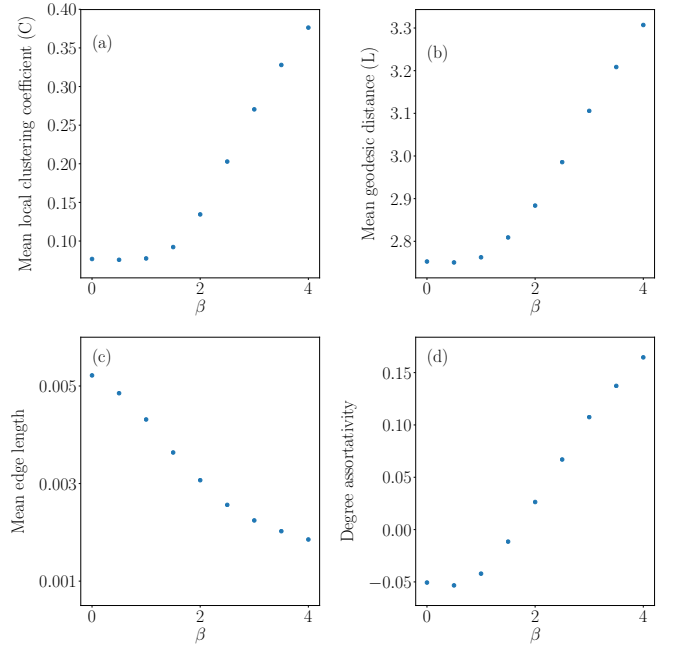

FIG. 7. Some characteristics of the networks from our SPA model for different values of the spatial decay parameter β for networks with $n = 10010$ nodes, $m = 5$ edges for each new node as we add it, and 10 instantiations of our model. We show computations of (a) mean local clustering coefficient, (b) mean geodesic distance, (c) mean edge length, and (d) degree assortativity.

(see Fig. 9) change as a function of the number of nodes for SPA networks. In Fig. 8, for spatial decay parameter values of $\beta \in \{0, 1, 2, 3, 4\}$, the mean geodesic distance seems to increase logarithmically scaling with the number n of nodes according to numerical simulations. This is consistent with the observations for a 1D SPA in [18], which found mean geodesic distance to have logarithmic scaling with respect to n based on numerical simulations.

In Fig. 9, we see for $\beta \in \{0, 1, 2\}$ that the mean local clustering coefficient decays sharply for progressively larger n, and it is possible that it may approach 0 as $n \to \infty$. Meanwhile, for $\beta = 4$ and $\beta = 3$, we do not observe such sharp decay, at least for the examined values of n. Our numerics suggest the possibility that it may be the case that there is a value β_c such that $\beta > \beta_c$, one obtains a positive mean local clustering coefficient in the limit that $n \to \infty$.

IV. A SPATIAL CONFIGURATION MODEL

We now generalize a configuration model [2, 26] to incorporate spatial considerations. Configuration models are among the most important random-graph models, as they are used frequently as reference models (including as null models in community detection [19, 20]) for a large
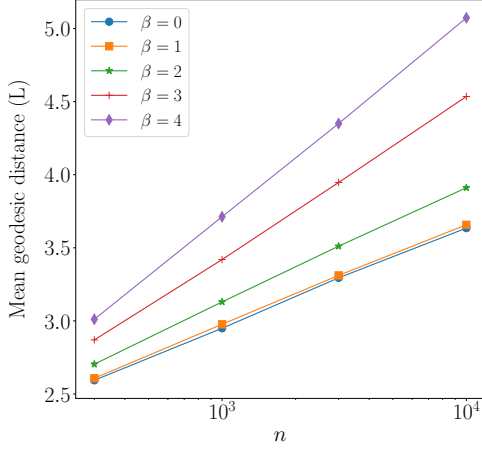
FIG. 8. Mean geodesic distance of our SPA networks as a function of the number of nodes in a network for several values of the spatial decay parameter β. We show the number n of nodes shown on a logarithmic scale. Each point in the plot represents a mean of the mean geodesic distances over 10 instantiations of our model for each value of n and β.
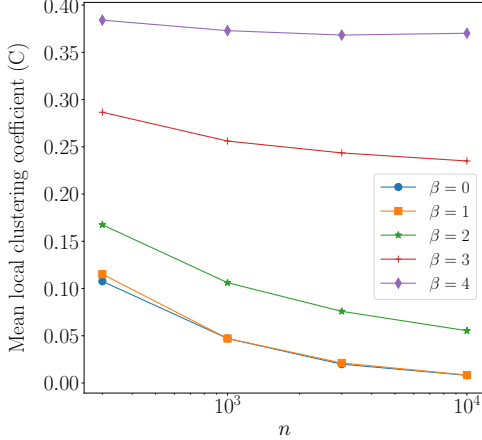

FIG. 9. Mean local clustering coefficients of our SPA networks as a function of the number of nodes in a network for several values of the spatial decay parameter β. We show the number n of nodes on a logarithmic scale. Each point in the plot represents a mean of the mean local clustering coefficients over 10 instantiations of our model for each value of n and β.

variety of network analyses. We envision that spatial analogs of configuration models will be similarly helpful for spatial networks.

A. Description of the model

Our goal in developing a spatial configuration model is to preserve the degree sequence of an input network, modifying the adjacency according to some rule that incorporates spatial embedding. Specifically, nodes are embedded in a latent space and assigned a number of edge stubs according to the degree sequence of the input network. Then, edge stubs are connected (in a process similar to a non-spatial configuration model [26]), but rather than selecting edge stubs uniformly at random, we preferentially connect stubs that are spatially close.

We must make specific choices for this process. We must choose how to assign node locations. They can be assigned uniformly at random, or according to a different randomization. If the input network is embedded in space and includes node locations, we can also consider leaving the node locations unmodified. We must also choose how to bias stub selection to connect spatially close stubs. In our model, we use a deterrence function in a similar fashion as the previous SPA and GF models. Our procedure is the following:

1. For each node, assign a new location, which we choose uniformly at random, in $[0,1] \times [0,1]$.

2. To each node v_l with degree k_l, assign k_l stubs to it. We denote the number of unmatched stubs for a node v_l at time step t by $u(v_l, t)$.

3. Choose a stub from step (2) uniformly at random. We label its associated node as v_i.

4. Choose a second stub with a probability proportional to $h(r_{i,j})$. That is, select a stub from node v_j with probability

$$p(v_i, v_j, t) = \frac{u(v_j, t)h(r_{i,j})}{\sum_{l \neq i} u(v_l, t)h(r_{i,l})} .$$

5. Connect the two stubs from steps (3) and (4) to each other with an undirected, unweighted edge.

6. Repeat steps (3)–(5) until all we have matched all stubs to form edges.

Because $r_{i,j}$ is independent of the time step, we can make the above process more efficient by calculating the pairwise Euclidean distance between all node pairs in a network (there are $O(n^2)$ such pairs to calculate) and store it for reuse in each stub-choosing step. After this, the algorithm takes $O(|E|n)$ time to run, where E denotes the set of edges and $|E|$ is the number of edges.

In formulating a spatial configuration model, one needs to decide whether to allow multi-edges and/or allowing self-edges. Since $h(r_{i,i}) = 0^{-\beta}$ does not have a well-defined value, we choose not to consider self-edges to get around this issue. We allow multi-edges.

There are also many other types of spatial configuration models that one can develop. For example, one can randomize the positions of nodes while preserving the adjacency matrix of the network.

As with non-spatial configuration models [26], choices regarding the implementation of the spatial configuration process should depend on the application and question of interest. One common application of a configuration model is to use it as a null model in community detection [19, 20, 37], where the exact choice of null model greatly affects which communities are detected.
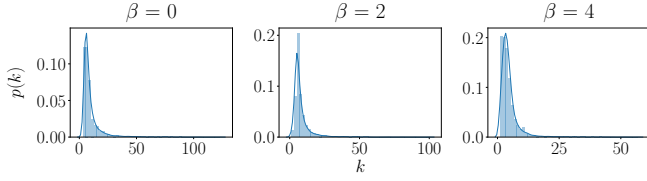
FIG. 10. Degree distributions for single instances of the spatial configuration model, with degree sequence given by a BA network with $n = 1010$ nodes and $m = 5$ new edges for each node that we add after the seed, for spatial decay parameters of $\beta = 0$, $\beta = 2$, and $\beta = 4$.
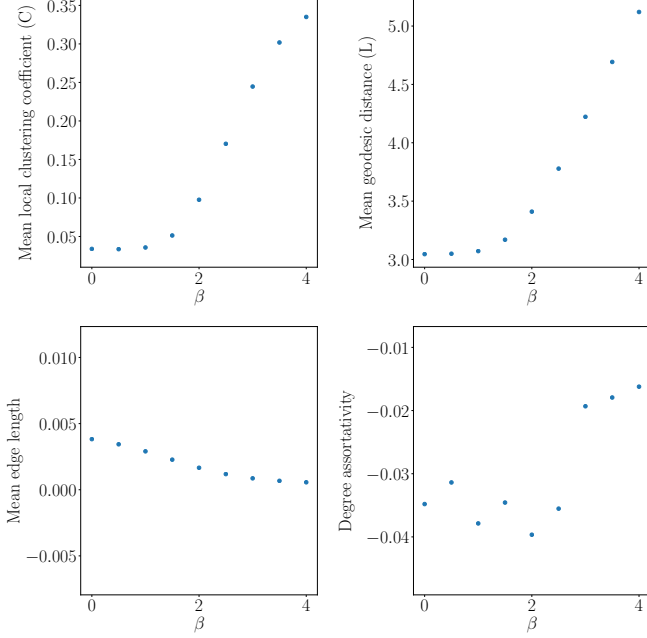

FIG. 11. Some characteristics of the networks from our spatial configuration-model networks for various values of β. We use a degree sequence given by a BA model with $n = 1010$ nodes and $m = 5$ new edges for each node that we add after the seed. A new BA network is generated for each instantiation. We take means of the network characteristics over 30 instantiations of our model. We show computations of (a) mean local clustering coefficient, (b) mean geodesic distance, (c) mean edge length, and (d) degree assortativity.

B. Computational results

To illustrate properties of our spatial configuration model, we start by generating a standard BA network. The seed network for the BA network is a 10-clique, each new node has $m = 5$ stubs, and we grow the network for a total of $T = 1000$ time steps (such that the final network has $n = 1010$ nodes). We then use the degree sequence from this network as the degree sequence for networks that we construct from our spatial configuration model. Note that a new BA network is generated for each application of the spatial configuration model.
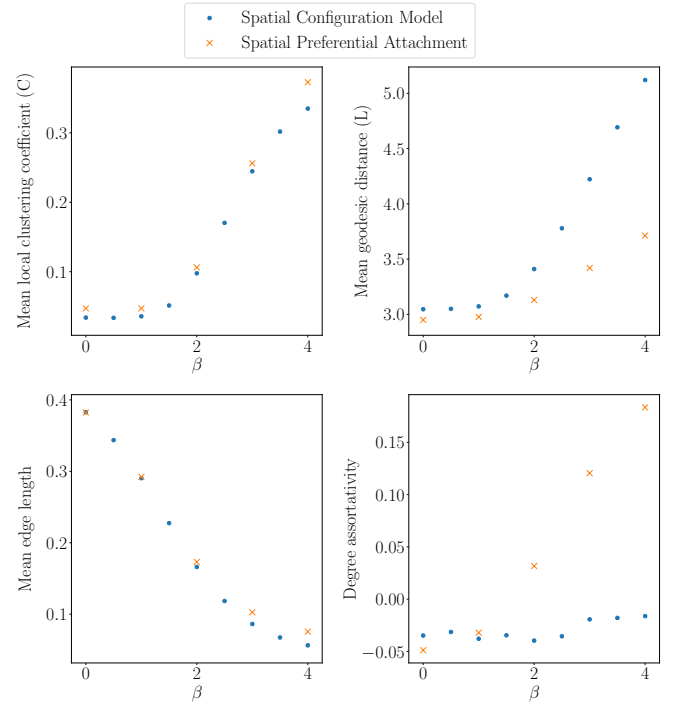
In our spatial configuration model, we observe (as in

FIG. 12. Comparison of the characteristics of our spatial configuration model networks (blue circles) with characteristics of SPA networks (orange crosses) with $n = 1010$ nodes and $m = 5$ new edges for each node that we add. The spatial configuration networks are the same networks (configurations of BA models) used to plot Fig. 11, while the SPA networks are the same as were generated in Section III.

the SPA and GF models) that as we increase the spatial decay parameter β, the mean local clustering coefficient and mean geodesic distance increase, whereas the mean edge length decreases. The degree assortativity of networks from our spatial configuration model does not appear to have a clear correlation with β, unlike with the SPA and GF models.

For comparison, we include an additional scatter plot of these characteristics for our spatial configuration model (for which we use degree sequences generated by BA networks), alongside scatter plots from our SPA network from Section III. They both use a 10-clique as the seed network, with the same number of nodes ($n = 1010$), and number of new edges per node that we add ($m = 5$).

V. SPATIAL STRENGTH CENTRALITY MEASURE

Our explorations of spatial network models raise an interesting question: Can we quantify the how strong the effects of the spatial embedding and choice of spatial decay function $h(r)$ are? We have observed that larger values of the spatial decay parameter β lead to more prominent spatial effects on network topology. However, it is

desirable to be more systematic about our analysis of spatial networks. For example, it is important to compare different choices of the function $h(r)$, different sizes and dimensions of the ambient space, and different distributions of nodes in space. Therefore, we define a centrality measure for spatial networks that we call *spatial strength centrality*, and compare it across several synthetic and empirical spatial networks.

A. Definition and description of spatial strength centrality

To develop a notion of centrality for spatial networks, we proceed as follows. Let $N(v_i)$ be the set of nodes to which node v_i is adjacent (i.e., its neighborhood). We then calculate the mean edge distance — which we take to be Euclidean for concreteness, but one can also consider other metrics — from node v_i to each other node in its neighborhood. For a node v_i with at least one incident edge, this distance is

$$L(v_i) = \frac{\sum_{v_j \in N(v_i)} d(v_i, v_j)}{k_i} \frac{1}{\langle L \rangle}, \qquad (11)$$

where

$$\langle L \rangle = \frac{\sum_{(v_i, v_j) \in E} d(v_i, v_j)}{n \langle k \rangle} \qquad (12)$$

and $d(i, j)$ is the Euclidean distance between nodes v_i and v_j. The left fraction in $L(v_i)$ gives the mean edge length of v_i; we then normalize it by $\langle L \rangle$, the mean edge length in the network. We now calculate the mean neighbor degree of each node v_i and normalize it by the mean degree in the network. That is,

$$K(v_i) = \frac{\sum_{v_j \in N(v_i)} k_j}{k_i} \frac{1}{\langle k \rangle}. \qquad (13)$$

We then combine the above two quantities to calculate the spatial strength centrality

$$S(v_i) := \frac{1}{L(v_i) K(v_i)} \qquad (14)$$

of each node v_i with at least one incident edge.

Our motivation in defining $S(v_i)$ is to capture a notion of whether nodes are adjacent to each other because they are spatially close or because they are adjacent to each other for a topological reason. Heuristically, we "reward" a node for being adjacent to nodes that are spatially close, and we "penalize" it for being adjacent to "hubs" (i.e., to nodes with large degree).

As an example, consider a network of flights between airports. Suppose that a small airport v_i is adjacent only to a hub v_j, which is also far away from it geographically. Node v_i has a small spatial strength $S(v_i)$ because its edge to v_j does not arise from it being nearby, but instead because v_j is a hub. This idea is captured by $S(v_i)$ because both $L(v_i)$ and $K(v_i)$ are large, as the one edge of v_i is a long edge and its neighbor has large degree. Therefore, from (14), we see that $S(v_i)$ is small.

Conversely, consider a granular network [38]. In this network, nodes are adjacent if they are touching (or at least sufficiently close to be construed as touching). Moreover, because of physical constraints, the number of edges that are attached to a node is limited to be a small number. Therefore, an arbitrary node v_j in this network has short edges, so $L(v_j)$ is small. Additionally, every node in the neighborhood of v_j has small degree, so $K(v_j)$ is small. Therefore, from (14), we see that $S(v_j)$ is larger in this example than $S(v_i)$ from the previous airport example.

It can also be informative to calculate a network's mean spatial strength centrality

$$\langle S \rangle = \frac{\sum_i S(v_i)}{n}. \qquad (15)$$

In the previous examples, for instance, we expect a granular network to have a larger value of mean spatial strength centrality than the air-traffic network, because the former's network topology has more stringent constraints from the system.

There are several important considerations for calculating mean spatial strength:

1. We have normalized all quantities, so we expect to be able to meaningfully compare the value of $\langle S \rangle$ for different types of networks, including ones with different size scales and spatial embeddings. However, $\langle S \rangle$ is unbounded (in particular, it is not confined to values between 0 and 1), so we need to be careful about interpreting its values and comparisons of these values.

2. The mean spatial strength $\langle S \rangle$ is nonnegative.

3. We normalized $K(v_i)$ by the mean degree $\langle k \rangle$, rather than by the largest degree $k_{\max}$ in a network, so it is not guaranteed to lie between 0 and 1.

4. Our formulas for $L(v_i)$ and $S(v_i)$ are not well-defined for nodes that have no edges, so we take these quantities to be 0 in these cases.

B. Computation of spatial strength on network models

As an initial test case, we consider the spatial strength centrality of the models that we defined earlier: the GF model, SPA model, and the spatial configuration model. We expect the mean spatial strength $\langle S \rangle$ to increase as we increase the spatial decay parameter β (see Fig. 13).

Interestingly, although there is generally a positive correlation between β and $\langle S \rangle$, it is not a linear relationship. The spatial configuration and SPA models yield S-shaped
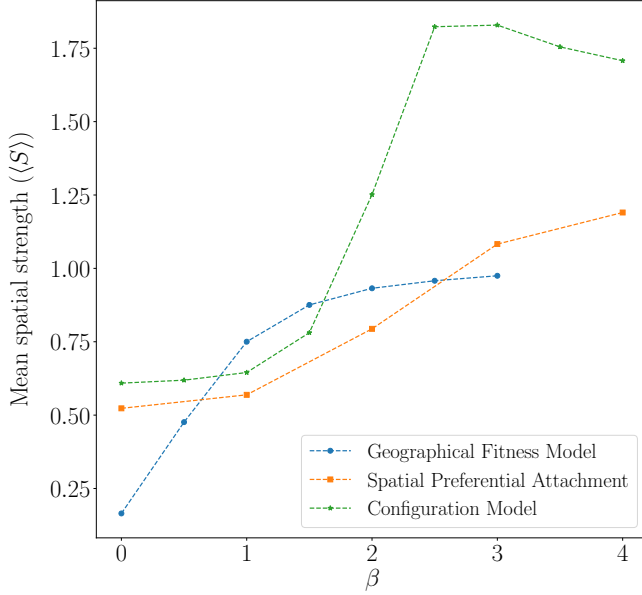
FIG. 13. Mean spatial strength for our geographical fitness model, our SPA model, and the spatial configuration model for various values of the spatial decay parameter β. For a given value of β, each point represents a mean over 20 realizations of the model. Our geographical-fitness-model networks have $n = 500$ nodes, and our SPA and spatial-configuration-model networks have $n = 1010$ nodes.
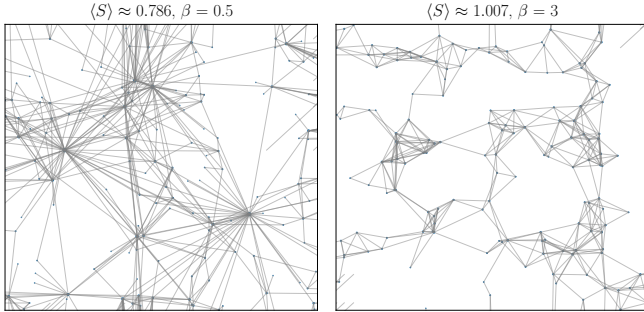

FIG. 14. Example networks from individual instantiations of the geographical fitness model with $n = 150$ nodes, which we place in 2D according to their assigned coordinates. We show examples with (left) $\beta = 0.5$ and $\langle S \rangle \approx 0.786$ and (right) $\beta = 3$ and $\langle S \rangle \approx 1.007$.

curves, and the GF model has an initially rapid increase of $\langle S \rangle$ with β before tapering off.

Contrary to expectations (which were for $\langle S \rangle$ to tend to increase with β), for the spatial configuration model, the mean spatial strength reaches a peak at about $\beta = 3$. For progressively larger β, the mean edge length of a network decreases, in turn decreasing the mean spatial strength (since $L(v_i)$ increases as $\langle L \rangle$ decreases according to (11), thus $S(v_i)$ decreases as $\langle L \rangle$ decreases). This sensitivity to mean edge length is a potential weakness in the centrality measure, and in Section V E we discuss potential ways to address this.
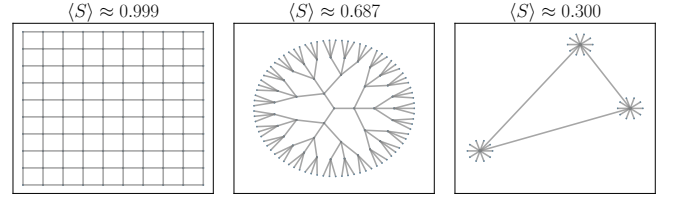

FIG. 15. Example networks with the indicated values for mean spatial strength. The depicted networks are (left) a lattice network, (center) a spatially-embedded Cayley tree, and (right) a hub-and-spoke example.

C. Examination of spatial strength on toy synthetic networks

In Fig. 14, we show individual instantiations of the GF model. To aid in developing intuition about the structure of networks that arise from this model, we also indicate their mean spatial strength centralities. To develop intuition about the centrality measure, we examine spatial strength centrality for several toy examples. We consider the following networks:

- **Square lattice.** The lattice network has x columns and y rows, giving it a total of $n = xy$ nodes. We space these nodes evenly in $[0, 1] \times [0, 1]$, and each node is adjacent to the nodes that are immediately north, south, east, and west of it. See the left plot of Fig. 15. When $x = y$, the mean spatial strength tends towards 1 as $n^2 \to \infty$.

- **Spatially-embedded Cayley tree.** This network has b branches and l layers. We start with one central node (layer 0). Inductively, every node in layer l is adjacent to b nodes in layer $l+1$, whose nodes are equally spaced in a circle of radius $l + 1$. See the center plot of Fig. 15.

- **Hub-and-spoke example.** This example has three "hub" nodes that are each spaced relatively far away from its 10 "spoke" nodes, which occur a circle around it. See the right plot of Fig. 15. This example demonstrates that a network with long-range connections can have a small mean spatial strength centrality.

- **Two-scale example.** Exploiting the definition of spatial strength centrality, we construct a network that consists of two pairs of nodes, where the nodes of each pair are adjacent to each (see Fig. 16). By making one edge short and the other arbitrarily large, the mean spatial strength centrality approaches infinity as the second edge becomes arbitrarily long. Therefore, we see that there exist networks with arbitrarily large values of mean spatial strength centrality.
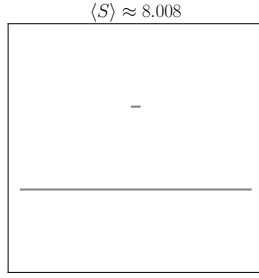
FIG. 16. This example of a two-scale network with two pairs of nodes illustrates that the mean spatial strength centrality $\langle S \rangle$ can be arbitrarily large.

D. Examination of spatial strength centrality on empirical and synthetic data sets

We now examine spatial strength centrality of several empirical networks, as well as on an RGG, which we defined in II. The data sets for fungal networks are from [11] (with 270 networks with between 68 and 2742 nodes and a mean of 819 nodes), and the data sets for city road networks are from [12] (with 101 networks with between 42 and 3871 nodes and a mean of 874 nodes). We show some example networks and their mean spatial strength centralities in Fig. 17.

map: check the numbers of networks in the data sets; for example, shouldn't the road networks have 100 data sets, not 101? I thought it was 20 cities from each of 5 continents, so what is the 101st network?

Andy: I double checked and there seemed to be 21 networks in the Africa folder, is this incorrect?

None of the model networks that we have explored thus far in the paper have yielded a mean spatial strength of greater than 2. However, many of the networks in both the city road and fungal data sets have a mean spatial strength centrality that is larger than 2. Therefore, there are structural features in spatial networks beyond the ones in the models in this paper.

We observed that in networks in the data sets with greatest $\langle S \rangle$, there were areas of space in which nodes are very close together. For example, in a city road network (see the left side of Fig. 17), multiple nodes (i.e., street intersections) can occur along curves in a road. Such nodes tend to have short edges between them, and these networks therefore include edges at multiple spatial scales. As we saw in the example in Fig. 16, having both very long edges and very short edges can yield a large mean spatial strength centrality.

In Fig. 18, we show a scatter plot of many of these networks and their mean spatial strength centralities.

E. Alternate formulations of spatial strength centrality

As we saw in the computations of spatial strength centrality, it sometimes captures some aspects of spatial embeddings influencing network topology, but it was not always successful at doing so (e.g., the fact that $\langle S \rangle$ is not an increasing function of β for the spatial-configuration-model networks). To explore these issues further, the following considerations are worth examining:

- **Calculating spatial strength centralities for edges.** In (14), we defined spatial strength centrality as a combination of mean neighbor degree and the mean edge length of a node. One might instead consider the "spatial strength" of edges instead of nodes. For example, for an edge (i, j) one could sum the degrees of v_i and v_j, and divide by the length of the edge.

- **Normalization of mean edge length.** We noted (see Section V B) that $\langle S \rangle$ is sensitive to mean edge length. This is because $L(v)$ increases and $S(v)$ decreases as $\langle L \rangle$ decreases. It may instead be useful to normalize $L(v)$ by the maximum pairwise edge length in a network, rather than by the mean edge length.

  The point of normalizing $L(v)$ by mean edge length was in order to be able to compare networks of different sizes. An alternate way to do this would be to normalize $L(v)$ by the geographic diameter of the network, as measured by the maximum pairwise distance between nodes of the network.

- **Comparison of a spatial network to null models.** The goal of calculations of spatial strength centrality is to determine how a spatial embedding affects network topology. To that end, instead of a centrality measure, it may be desirable compare a given network to a spatial null model to determine how a spatial embedding affects the adjacency matrix (i.e., topology) of a network. For such purposes, the spatial configuration model may be useful as a null model.

VI. CONCLUSIONS AND DISCUSSION

We have developed and examined a straightforward method for generalizing generative models of networks to incorporate spatial information by using a deterrence function $h(r)$ that decays with the distance r between nodes to adjust the probability that an edge occurs between a pair of nodes. For concreteness, we used a Euclidean distance and the power-law decay rate $h(r) = r^{-\beta}$, but our formulation allows one to make diverse choices of both metric and decay function.

One illustration of how to augment existing network models, rather than define new spatial network models

$$\langle S \rangle \approx 2.339 \qquad \langle S \rangle \approx 2.279 \qquad \langle S \rangle \approx 1.228$$
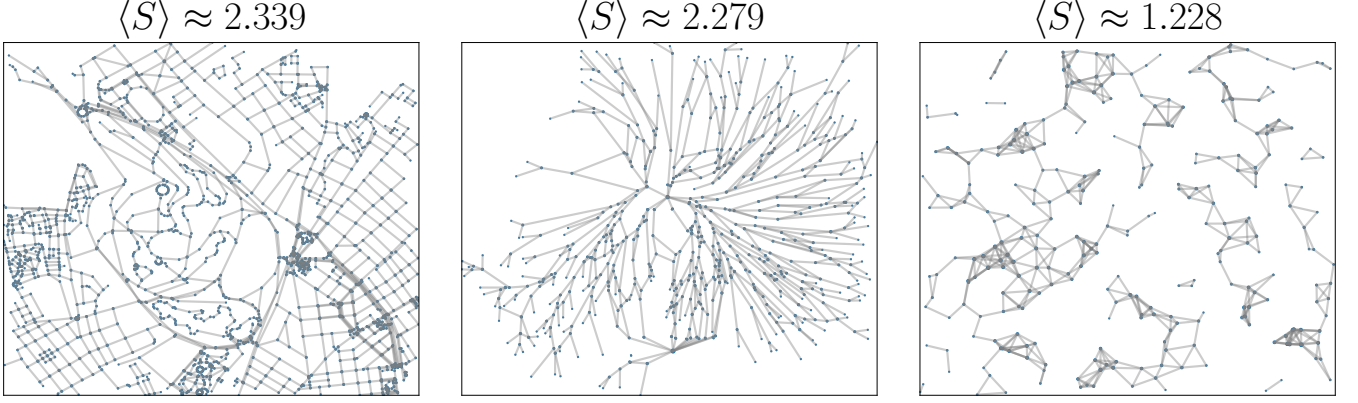

FIG. 17. Example networks and mean spatial strength centralities empirical networks and a random geometric graph. We show (left) a road network from Tunis, Africa with 1731 nodes; (center) a fungal network labeled "Pv_M_I+4R_U_N_21d_4" (see [11]) with 641 nodes; and (right) a random geometric graph with $n = 500$ nodes and distance parameter $r_c = 0.07$.
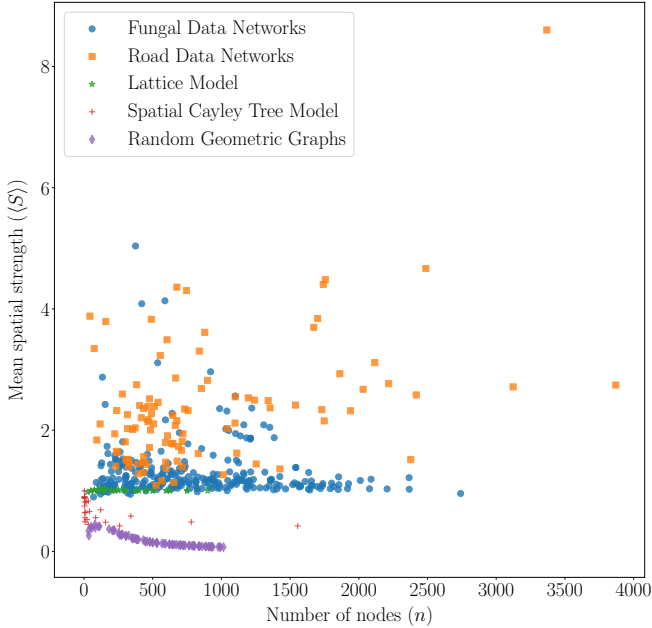

FIG. 18. Scatter plot of mean spatial strength versus the number of nodes in several synthetic and empirical networks. We use fungal network data from [11], road network data from [12], and RGGs with a distance parameter of $r_c = 0.07$. The spatial Cayley trees in this plot range from 1 to 6 branches, and their depths range from 1 to 4. The lattice networks have 5, 10, 15, 20, 25, or 30 rows and columns (and any combination of rows and columns of those sizes); on the plot, the form an almost horizontal line at a mean spatial strength centrality of about 1.

from scratch, is with our formulation of a spatial configuration model. We also extended a GF model with $h(r)$, and examined an SPA model that used this deterrence function. We examined the properties of these models and compared them to random geometric graphs and empirical spatial networks from two disparate applications. To examine the structure of spatial networks more deeply — and, in particular, to try to separate the effects of spatial embeddings and other influences on network architecture, we defined a spatial strength centrality, which allowed us to estimate how strongly the effects of a network's ambient space (in which its nodes are embedded) affects observed network topology. We examined spatial strength centrality on several toy networks and compared it on a diverse set of synthetic and empirical spatial networks.

Spatial networks have diverse uses in the modeling of networks from empirical data, and the models that we have examined in this paper should help in such efforts. We anticipate that further exploration of spatial null models (e.g., using our spatial configuration model and generalizations of it) will be particularly insightful, as they provide baselines for comparisons with empirical data. To explore the diverse effects of spatial embeddings (and other effects of space) on network topology, it is also important to further analyze deterrence functions $h(r)$ and a variety of notions of spatial strength centrality.

[1] M. Barthelemy, *Morphogenesis of Spatial Networks* (Springer International Publishing, Cham, Switzerland, 2018).

[2] M. E. J. Newman, *Networks*, 2nd ed. (Oxford University Press, Inc., Oxford, UK, 2018).

[3] P. D. Hoff, A. E. Raftery, and M. S. Handcock., Journal of the American Statistical Association **97**, 1090 (2002).

[4] R. J. Williams and N. D. Martinez, Nature **404**, 180 (2000).

[5] P. Crucitti, V. Latora, and S. Porta, Phys. Rev. E **73**, 036 (2006).

[6] A. Barrat, M. Barthelemy, and A. Vespignani, Journal of Statistical Mechanics: Theory and Experiment **2005** (2005).

[7] D. J. Aldous and J. Shun, Statist. Sci. **25**, 275 (2010).

[8] M. Bogu, R. Pastor-Satorras, A. Daz-Guilera, and A. Arenas, Phys. Rev. E **70**, 056 (2004).

[9] F. Klimm, D. S. Bassett, J. M. Carlson, and P. J. Mucha, PLoS Comput Biol **10** (2014).

[10] D. A. Fair, A. L. Cohen, J. D. Power, N. U. F. Dosenbach, J. A. Church, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen, PLoS Comput Biol **5** (2009).

[11] S. H. Lee, M. D. Fricker, and M. A. Porter, Journal of Complex Networks **5**, 145 (2017).

[12] S. H. Lee and P. Holme, Physical Review Letters **108**, 128 (2012).

[13] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and MarcelloTomasini, Physics Reports **734**, 1 (2018).

[14] M. T. Gastner and M. E. J. Newman, J. Stat. Mech. **2006** (2006).

[15] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, UK, 2003).

[16] N. Masuda, H. Miwa, and N. Konno, Physical Review E **71**, 036108 (2005).

[17] J. P. Taylor-King, D. Basanta, S. J. Chapman, and M. A. Porter, Physical Review E **96** (2017).

[18] R. Xulvi-Brunet and I. M. Sokolov, Physical Review E **66** (2002).

[19] M. Sarzynska, E. A. Leicht, G. Chowell, and M. A. Porter, Journal of Complex Networks **4**, 363 (2016).

[20] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, PNAS **108**, 7663 (2011).

[21] K. Zuev, M. Boguña, G. Bianconi, and D. Krioukov, Scientific Reports **5** (2015).

[22] M. Bradonji, A. Hagberg, and A. G. Percus, Internet Mathematics **5**, 113 (2011).

[23] Y. Ide, N. Konno, and N. Obata, Internet Mathematics **6**, 173 (2010).

[24] S. S. Manna and P. Sen, Physical Review E **66** (2002).

[25] S.-H. Yook, H. Jeong, and A.-L. Barabsi, PNAS **99** (2002).

[26] B. K. Fosdick, D. B. Larremore, J. Nishimura, and J. Ugander, SIAM Review **60**, 315 (2018).

[27] G. Caldarelli, A. Capocci, P. D. L. Rios, and M. A. Munoz, Physical Review Letters **89**, 258 (2002).

[28] N. Masuda, H. Miwa, and N. Konno, Physical Review E **70**, 036124 (2004).

[29] M. Boguña and R. Pastor-Satorras, Physical Review E **68**, 036112 (2003).

[30] S. A. Frank, Journal of Evolutionary Biolology **22**, 1563 (2009).

[31] J. Dall and M. Christensen, Physical Review E **66**, 016121 (2002).

[32] A. Antonioni and M. Tomassini, Physical Review E **86** (2012).

[33] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[34] L. Ferretti and M. Cortelezzi, Physical Review E **84** (2011).

[35] W. Aiello, A. Bonato, C. Cooper, J. Janssen, and P. Pralat, Internet Mathematics **5**, 175 (2008).

[36] J. Emmanuel and P. Mörters, The Annals of Applied Probability **25** (2015).

[37] D. S. Bassett, E. T. Owens, M. A. Porter, M. L. Manning, and K. E. Daniels, Soft Matter **11**, 2731 (2015).

[38] L. Papadopoulos, M. A. Porter, K. E. Daniels, and D. S. Bassett, Journal of Complex Networks **6**, 485 (2018).