TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Qiu, Z., Balogh, G. S., Liu, X., Van Mieghem, P. F. A., & Kitsak, M. A. (2026). Geometric organization and inference of shortest path nodes in soft random geometric graphs. *Physical Review Research*, *8*(3), Article 033149. https://doi.org/10.1103/dq2j-z8k6

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Geometric organization and inference of shortest path nodes in soft random geometric graphs

Zhihao Qiu [1], Sámuel G. Balogh [1,2], Xinhan Liu [1], Piet Van Mieghem [1] and Maksim Kitsak [1,3,*]

[1] *Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands*

[2] *National Laboratory of Health Security, HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary*

[3] *Department of Informatics, Indiana University Bloomington, Bloomington, Indiana, USA*

(Received 9 February 2026; accepted 17 June 2026; published 7 August 2026)

The shortest path problem is related to many dynamic processes on networks, ranging from routing in communication networks to signaling in molecular interaction networks. When the network is fully known, the shortest path problem can be solved precisely and in polynomial time. If, however, the network of interest is only partially observable, the shortest path problem is no longer straightforward. Inspired by the shortest path problem in partially observable networks, we investigate the geometric properties of shortest paths in *Euclidean* soft random geometric graphs (SRGGs). We find that shortest paths are aligned along geodesic curves connecting shortest path end points. The strength of the shortest path alignment, as quantified by the average distance to geodesic from shortest path nodes and the average path stretch, is higher for larger SRGGs with short-range connections. In addition, we find that the strength of the shortest path alignment is nonmonotonic with respect to the average degree of the SRGG. Based on these observations, we establish the conditions under which the alignment of shortest paths may be sufficiently strong to allow the identification of shortest path nodes based on their proximity to geodesic curves. We show that in partially observable networks with uncertain node positions, our geometric approach can outperform network-based shortest path algorithms. In practical settings, our findings may have applications to navigation, wireless routing, and flow characterization in infrastructure networks.

DOI: 10.1103/dq2j-z8k6

I. INTRODUCTION

In our work, we study the geometric organization and inference of shortest path nodes in soft random geometric graphs (SRGGs).

The shortest path $\mathcal{P}_{ij}^{*}$ from a node i to a node j in a network G is an ordered sequence of distinct nodes to traverse from i to j, such that the sum of the link weights along the path is minimized. The shortest path has been extensively studied in the literature due to its relevance to a large array of real problems in routing [1–5], transportation [6–10], spreading processes [11–15], and search [16,17].

Conventional shortest path-finding algorithms like Dijkstra [18,19] and Bellman-Ford [19–22] find shortest paths precisely and in polynomial time if the network of interest is fully observable. However, the accuracy of conventional methods quickly decreases as the number of missing network links increases [16], due to the extreme sensitivity of shortest paths: Even a single missing (false negative) or a spurious (false positive) link may change the path entirely. This is the case since conventional shortest path algorithms rely on the iterative exploration of network topology and are very likely to fail if some of the links or nodes are not known [16]. Unfortunately, large-scale social, technological, and biological networks are often *substantially* incomplete: The majority of links and/or nodes are not known [23–25].

Finding shortest paths in partially observable networks using network reconstruction methods is very difficult, if not impossible. The most popular network reconstruction task is link prediction. Link prediction literature is dominated by *ad hoc* methods ranking missing link candidates [26]. These approaches allow one to identify the most likely individual missing link candidates, but cannot quantify the joint likelihood that several links are missing simultaneously. More elaborate network reconstruction methods, e.g., those based on the Bayesian formalism [27], allow one to formulate the probability of observing any subgraph, including a path, in a network. In practice, however, this probability depends on the choice of the null model and must be computed numerically, which is often computationally expensive.

Random matrix theory allows one to denoise real networks of interest, and notable advances exist in community detection [28–30], denoising correlated networks [31,32], and studies of centrality measures [33]. Random matrix theory in application to partially observed networks can, in principle, *denoise* networks by treating missing or spurious information as noise. However, it remains an open question whether shortest path nodes can be accurately identified after network denoising.

*Contact author: maksim.kitsak@gmail.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Our work is motivated by the recent finding that proposed a method for finding shortest and nearly shortest path nodes, making use of their geometric alignment in hyperbolic embeddings of these networks: Shortest path nodes are likely found in the geometric vicinity of geodesic curves connecting shortest path end points [16]. To find a shortest path between nodes A and B in network G, one needs to obtain a hyperbolic representation of the network first, i.e., map network nodes to points in a hyperbolic disk. Node coordinates obtained as a result of hyperbolic embeddings allow one to draw a geodesic curve $\gamma(A, B)$ connecting shortest path end points A and B. The likelihood for a node C to belong to a shortest path between A and B can be estimated using the distance to geodesic $d(C, \gamma(A, B))$ from C to $\gamma(A, B)$, the smaller the distance the higher the likelihood.

Kitsak *et al.* [16] have demonstrated that hyperbolic embeddings can be used to find shortest paths even in extremely incomplete networks, with as few as 10% of network links present, leading to promising applications in biomedicine and communication networks. The caveat of the method is that it is limited to networks whose organization is consistent with hyperbolic geometry. It is generally accepted that networks with scale-free degree distributions and strong clustering coefficients are best embedded into hyperbolic spaces [34].

While hyperbolic embeddings are effective for networks with strong degree heterogeneity and inherent hierarchical organization, features that facilitate alignment between shortest paths and geodesics, these properties are entirely absent in Euclidean geometric networks. Such networks are more degree homogeneous and it remains unclear whether geometry alone can still provide equally meaningful information about shortest path nodes. Motivated by this, we investigate whether shortest paths in *Euclidean* geometric networks are sufficiently aligned and identifiable with geometric methods. On the one hand, our work is fueled by earlier findings in *Euclidean* random geometric graphs (RGGs) demonstrating a certain degree of alignment of shortest paths [35–39]. On the other hand, our work is inspired by the observation that many networks of interest are embedded into *Euclidean* spaces. Indeed, nodes of wireless communication networks are base stations placed in various geographic areas, and the quality of communication between any two base stations depends on the *Euclidean* geographic distance between them [40–43]. Another example is transportation networks, where nodes are intersections and road segments connecting them are links [44,45]. Finally, of relevance to this study are critical infrastructure networks, e.g., power grids consisting of power stations connected with transmission lines [46,47]. In all these networks, nodes are characterized by their geographic positions, and links are typically connecting nearby nodes.

Our work lies at the intersection of three actively studied areas: random graph models, network geometry, and statistical inference. Therefore, before proceeding with results, we position our work with respect to existing findings.

Shortest paths in random networks have been extensively studied in graph theory, network science, and computer science due to their relevance to a large number of practical problems. While there is a large literature corpus on shortest paths and routing paths in random graphs (see, e.g., Refs. [4,48,49]), there is limited literature on shortest paths in partially observed networks and geometric properties of shortest paths in geometric networks.

The best-studied geometric network model is the RGG. However, these studies tend to focus not on shortest paths but on RGG connectivity and percolation properties [50]. Recently, random hyperbolic graph (RHG) emerged as a popular network model due to its ability to reproduce common topological properties of real networks, such as sparseness, scale-free degree distributions, strong clustering [34,51], self-similarity [52], and high modularity [53–56].

In random graphs built in hyperbolic spaces, the efficiency of greedy routing has received substantial interest. In contrast to conventional routing, greedy routing relies on forwarding data packets using node coordinates, such that the packet is forwarded to the node closest to the destination in the latent space. The efficiency of greedy routing in RHG has been conjectured to be due to greedy routing paths closely aligned along geodesic curves in hyperbolic space [51]. However, a recent work challenged this conjecture by demonstrating that topological properties of hyperbolic networks are not strictly congruent with the geometry of underlying latent spaces [57]. Almost concurrent with this work, it has been shown that shortest paths in RHG and in real networks embedded in hyperbolic spaces are, nevertheless, sufficiently aligned along geodesic curves and shortest path nodes can be found using the distance to geodesic [16]. While shortest paths are intuitively expected to display a certain degree of geometric organization in *Euclidean* spaces, existing works tend to focus either on upper bounds of shortest path lengths [36–39], on fluctuations and geometric properties of path shapes [35,58], or on robust and data-driven shortest path optimization under uncertainty [59].

In summary, shortest paths in RHG and networks embeddable in hyperbolic spaces are aligned along geodesic curves, and shortest path nodes can be found even if a large fraction of network links are not observed. In *Euclidean* random network models, shortest paths are *expected* to align along geodesics, but this problem has not been studied directly. Further, it is not known if shortest path nodes in *Euclidean* random network can be identified using geometry of latent spaces. In our work, we fill in this knowledge gap by studying the extent of alignment of shortest paths in *Euclidean* SRGGs and develop a geometric method to find shortest path nodes in SRGGs under uncertainty of node coordinates. Soft random geometric graphs [35,60,61] can be regarded as a special case of the geometric inhomogeneous random graph (GIRG) model [62,63]. In an SRGG, nodes are positioned at random in a latent metric space $\mathcal{M}$ and connections between nodes are established independently with probabilities that are functions of distances between nodes in $\mathcal{M}$.

We find that in SRGGs, shortest path nodes tend to be significantly close to geodesic curves connecting path end points than expected by chance. In other words, we establish that shortest paths are geometrically aligned along geodesic curves. These alignments exhibit nontrivial behavior as a function of graph properties such as the average degree and the clustering coefficient. We find that under optimal conditions, the alignment of shortest paths allows one to find them using node coordinates in the latent space. Further, we

demonstrate that the geometric path-finding methods have a competitive advantage over conventional path-finding methods in situations when network topology is not completely observable.

The remainder of this paper is organized as follows. In Sec. II, we formalize the problem and introduce the SRGG model, together with the simulation framework used to quantify shortest path alignment and to infer shortest path nodes based on their distance-to-geodesic curves. In Sec. III A, we study the alignment of shortest paths in soft random geometric graphs as a function of their parameters. Relying on the results of Sec. III A, we examine the problem of the identification of shortest path nodes in Sec. III B. We conclude our work with the discussion and outlook in Sec. IV.

II. METHODS

A. Random geometric graph and soft random geometric graph

An RGG [50,64] is an undirected graph constructed by randomly placing N nodes into a metric space $\mathcal{M}$ and connecting any two nodes i and j if the distance d_{ij} between them is less than a certain radius r. In our work, we consider RGGs built in a two-dimensional unit square, where N nodes are placed uniformly at random. If the connection radius $r \ll 1$, the link density in the RGG with L links

$$p_{\mathrm{RGG}} \equiv \frac{L}{L_{\max}} \approx \pi r^2, \tag{1}$$

where $L_{\max} = \frac{1}{2}N(N-1)$ (see, e.g., Ref. [65]).

Thus, ignoring boundary effects, the probability that a node i has exactly k connections is

$$\Pr[D_i = k] = \binom{N-1}{k} (p_{\mathrm{RGG}})^k (1 - p_{\mathrm{RGG}})^{N-k-1}, \tag{2}$$

and the expected degree of a node is

$$\mathbb{E}[D_i] = (N-1)p_{\mathrm{RGG}}, \tag{3}$$

where p_{RGG} is the link density given by Eq. (1).

An SRGG can be regarded as a generalization of the random geometric graph in which links are established independently with probabilities that are functions of distances between the nodes in $\mathcal{M}$. An SRGG can be generated as follows:

(1) Sprinkle N nodes with probability density function (pdf) $\rho(x)$ into latent space $\mathcal{M}$.

(2) Calculate the distances $d_{\mathcal{M}}(ij)$ between each (i, j) node pair.

(3) Connect each (i, j) node pair independently with probability p_{ij}, which is a function of distances $d_{\mathcal{M}}(ij)$ between nodes i and j, $p_{ij} = f(d_{\mathcal{M}}(ij))$.

While, in principle, $f : R^+ \to [0, 1]$ can be any function, f is often selected to be a decreasing function to model networks where connections over short distances are preferable. One common choice for f is

$$f(d) = \beta e^{-\left(\frac{d}{d_0}\right)^{\eta}}, \tag{4}$$

where $\beta \in (0, 1]$, $d_0 > 0$, and $\eta > 0$ are tunable parameters. In particular, the $\eta = 1$ case is known as the Waxman graph model [50,66], and $\eta = \infty$ case reduces the SRGG model to the RGG. The choice of $\eta \in [2, 6.5]$ corresponds to the Rayleigh fading connection function [41,67] used to model radio signal's amplitude and phase fluctuations in multipath environments.

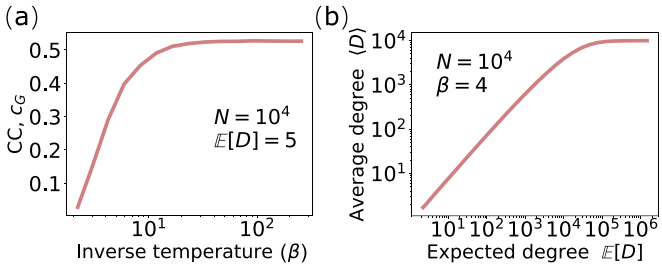


FIG. 1. Panel (a) shows an example of how the clustering coefficient changes with the input inverse temperature parameter in SRGGs. Panel (b) shows an example of how the real average degree changes with the input expected degree in SRGGs. The x axis is the input expected degree, while the y axis is the real average degree.

In this work, we study SRGGs with a different connection probability function, which is inspired by the Fermi-Dirac statistics:

$$f(d) = \frac{1}{1 + \left(\frac{d}{d_0}\right)^{\beta}}, \tag{5}$$

where $d_0 > 0$ and $\beta > 0$ are again parameters tuning topological properties of resulting graphs. The SRGG model is closely related to the $\mathbb{S}^D$ heterogeneous graph model [51] if node hidden variables in $\mathbb{S}^D$ are constant, and is also closely related to the GIRG model [68].

An SRGG model with assigned node coordinates can be interpreted through an analogy with a statistical-physics system in which energy levels ϵ_{ij} depend on node distances d_{ij} [51]. In this picture, links correspond to fermionic occupation of states, with at most one link per state [51]. Here, the parameter β acts as an inverse temperature, $\beta = 1/\tau$, and controls the average link length in the system; high values of β favor short-distance connections, while low values allow longer-distance connections. Following this statistical physics analogy, we refer to β as the inverse temperature parameter throughout the text.

We consider the case of SRGGs built on a unit square, with uniform node density, $\rho(\mathbf{x}) = 1$. Ignoring the boundary effects, one can establish that the expected degree of the SRGG model is given by

$$\mathbb{E}[D] = \mathbb{E}[D_i] \approx \frac{cNd_0^2}{\beta \sin\frac{2\pi}{\beta}}, \tag{6}$$

when $\beta > 2$, and the proportionality coefficient $c > 0$, arising due to finite-size effects, can be tuned numerically. Equation (6) indicates that parameter d_0 can be used to tune the expected degree of SRGG graphs. Indeed, greater d_0 values effectively rescale distances in the SRGG, allowing nodes to connect to other nodes over larger distances. Equation (6) holds well for small and moderate expected degree values [Fig. 1(b)]. For large expected degree values, the boundaries of the unit square become non-negligible, making Eq. (6) less precise.

In each simulation of the SRGG model defined by Eq. (5), the parameter d_0 is computed from Eq. (6) as $d_0 = \sqrt{\frac{\mathbb{E}[D]\beta \sin(\frac{2\pi}{\beta})}{cN}}$. To fix the remaining constant c, we first introduce an arbitrarily chosen reference configuration, in our case $N = 10^4$, $\beta = 128$, and $\mathbb{E}[D] = 1$. For this setup, we determine c numerically such that the theoretical expectation $\mathbb{E}[D]$ approximately matches the realized average degree of the generated network yielding $c \approx 25$.

Importantly, due to finite-size effects and the intrinsic stochasticity of the SRGG construction, the theoretically expected degree generally does not exactly coincide with the empirically observed average degree for all β, N unless c is properly calibrated. To make this distinction explicit, we carefully distinguish the theoretical expected degree $\mathbb{E}[D]$ of the SRGG model from the realized average degree $\langle D \rangle$ of the SRGG, which we measure as $\langle D \rangle \equiv 2L/N$ [see Fig. 1(b)]. Finally, for given values of N, β, and target average degree $\langle D \rangle$, the parameter c is fixed to its reference-configuration value, while $\mathbb{E}[D]$ is iteratively calibrated so that the resulting network exhibits an average degree equal to the prescribed $\langle D \rangle$ within a specified tolerance range.

While inverse temperature β also affects the average degree, it plays a far more important role by tuning the effective range of connections. When β is large, connections are mostly limited to distances $d \leqslant d_0$. Smaller β values increase the probability of connections at larger distances. As a result, β allows one to tune the average clustering coefficient in SRGGs [see Fig. 1(a)].

In our work, we parametrize SRGGs by the inverse temperature parameter β and expected degree $\mathbb{E}[D]$. To generate an SRGG with desired parameters β and $\mathbb{E}[D]$, we use Eq. (6) to determine parameter d_0. We then use parameters d_0 and β to determine connection probabilities for all node pairs in the network.

B. Alignment of shortest paths along geodesic curves

A geodesic $\gamma(i, j)$ is the shortest length curve connecting points i and j in space $\mathcal{M}$. The length of the geodesic $\gamma(i, j)$ is the distance d_{ij} between points i and j. In the case of *Euclidean* space, the geodesic $\gamma(i, j)$ is a straight line between i and j, while the distance d_{ij} represents the corresponding *Euclidean* distance.

The distance $d(q, \gamma(i, j))$ from a point q to the geodesic $\gamma(i, j)$ is the distance from point q to point $x \in \gamma(i, j)$ on the geodesic, such that distance $d(q, x)$ is minimized:

$$d(q, \gamma(i, j)) \equiv \min d(q, x), \tag{7}$$

$$\text{such that } x \in \gamma(i, j). \tag{8}$$

In *Euclidean* spaces, the distance from the point to a geodesic $d(q, \gamma(i, j))$ is the length of the perpendicular line segment from q to the geodesic line $\gamma(i, j)$.

The hop count h_{ij} of a path from a node i to a node j is the number of links of that path from node i to node j. Further, we define the path stretch S_{ij} as the total geometric distance accumulated along the path $\mathcal{P}^*_{ij}$:

$$S_{ij} \equiv \sum_{l \in \mathcal{P}^*_{ij}} d_l, \tag{9}$$

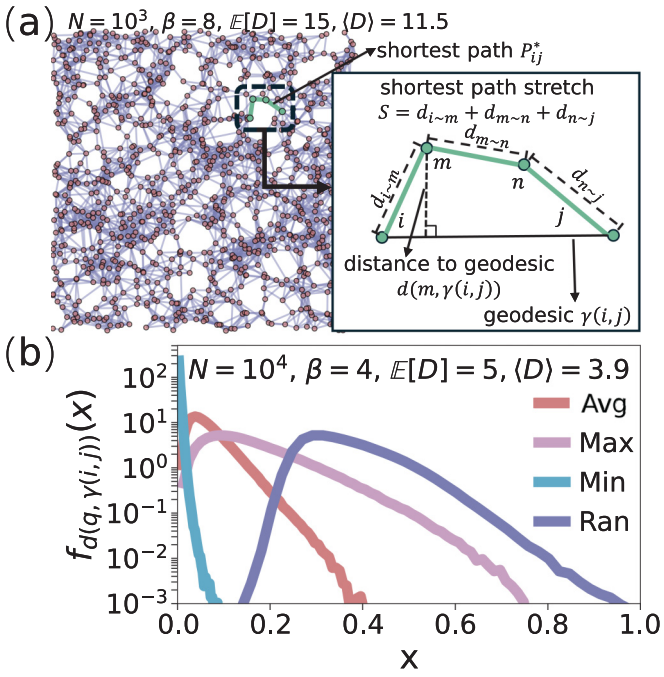


FIG. 2. (a) A toy example of a geodesic, distance from a node to the geodesic and a shortest path (highlighted by green) with the corresponding shortest path stretch indicated by the black dashed line along the green path. (b) The distribution of the average (avg), maximum (max), and minimum (min) distances between shortest path nodes and geodesics connecting shortest path end points. We compare the resulting distributions to (ran) those between randomly chosen nodes and the geodesic. Within an SRGG G, we randomly select 10^6 shortest path end point pairs and for each (i, j) pair, we find shortest path set $\mathcal{P}^*_{ij}$, geodesic $\gamma(i, j)$. For shortest path nodes in $\mathcal{P}^*_{ij}$, we compute the minimum (min), maximum (max), and average (avg) distance to the geodesic. In addition, for each shortest path set $\mathcal{P}^*_{ij}$, we created a random set $\mathcal{P}^{\text{rand}}_{ij}$ of the same cardinality containing randomly selected nodes. Using $\mathcal{P}^{\text{rand}}_{ij}$, we computed average distances to the geodesic (ran). All parameters used in the experiments are indicated in the respective panels.

where d_l is the *Euclidean* distance between the two end nodes of link l. Figure 2(a) depicts an example of a shortest path, a geodesic, the distance from a node on the path to the geodesic, and the shortest path stretch in an SRGG.

In an unweighted SRGG, there can be more than one shortest path $\mathcal{P}_{ij}$ between a node pair (i, j). The set $W_{P^*_{ij}}$ of all nodes constituting shortest paths between i and j are called "shortest path nodes," and $|W_{P^*_{ij}}|$ is the cardinality, i.e., the number of elements in the set.

We quantify the alignment of shortest paths along geodesic curves with two complementary methods. The first method is to compare distances from shortest path nodes to the geodesic with those from randomly selected nodes. The second method is to compute the shortest path stretch, i.e., the distance accumulated along shortest path links. The smaller the stretch the stronger the alignment [Fig. 2(a)].

While the distance-to-geodesic metric measures the physical deviation from the ideal straight route, it does not account for the additional distance incurred by traversing jagged, discrete edges. In contrast, the average stretch reflects the total

geometric length accumulated along the path, yet does not distinguish whether this arises from a single large spatial detour or from local zigzagging. Together, these two diagnostics provide a more complete two-dimensional view of the shortest path alignment phenomenon.

To assess the extent of the geometric alignment of shortest path nodes in SRGGs with given parameters, we design the following experiments:

(1) Generate an SRGG with the network size N, the expected degree $\mathbb{E}[D]$, and the inverse temperature parameter β. We then select a number of node pairs at random.

(2) For each node pair (i, j), we obtain all shortest path nodes $W_{\mathcal{P}^*_{ij}}$.

(3) Find the geodesic curve γ_{ij} connecting node i and j.

(4) Compute the distance to geodesic $d(q, \gamma(i, j))$ for all shortest path nodes $q \in W_{\mathcal{P}^*_{ij}}$ except for nodes i and j. Record the average, maximum, and minimum distance for all shortest path nodes.

(5) Randomly select $m \equiv |W_{\mathcal{P}^*_{ij}}| - 2$ nodes from the graph and compute the distance to the geodesic for each node.

(6) If there is a single shortest path connecting i and j, compute its stretch. If there are multiple shortest paths connecting i and j, select one shortest path at random and compute its stretch.

C. Predicting shortest path nodes using distance to geodesic

Kitsak *et al.* [16] proposed a method to infer shortest path nodes by measuring node proximities to geodesic curves connecting shortest path end points in hyperbolic embeddings of networks. We extend this approach to *Euclidean* SRGGs. We refer to this method as the "distance to the geodesic."

To infer shortest path nodes connecting nodes i and j of an SRGG in latent space $\mathcal{M}$, we first find the geodesic $\gamma(i, j)$ connecting the nodes in $\mathcal{M}$. Then, for each network node q excluding i and j, we find the distance to the geodesic $d(q, \gamma(i, j))$, and rank all network nodes in increasing order of distance to the geodesic. The closest to the geodesic nodes are then viewed as likely candidates for the shortest path nodes.

As a reference for comparison, we also infer shortest path nodes by first reconstructing network links and then identifying shortest path nodes using the Dijkstra algorithm. We consider two reconstruction strategies, denoted as SRGG+net and RGG+net.

In the SRGG+net experiment, we use known parameters β and $\mathbb{E}[D]$ and node coordinates $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ to rebuild 10^3 random graph instances G_ℓ, $\ell = 1, \ldots, 10^3$. For each instance ℓ, we find the set of shortest path nodes $W_{\mathcal{P}^*_{ij}(\ell)}$ using the Dijkstra algorithm on graph G_ℓ. As a result, shortest path node candidates are nodes that appear most frequently in all $W_{\mathcal{P}^*_{ij}(\ell)}$ sets.

The RGG+net experiment is a special case of the SRGG+net experiment when $\beta = \infty$. Since $\beta = \infty$ case corresponds to connecting nodes to geometrically closest nodes, the network reconstruction procedure is equivalent to the reconstruction of random geometric graphs, explaining the name of the experiment. For a specific network size N, expected degree $\mathbb{E}[D]$, and node coordinates $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$, there exists a unique RGG realization. Hence, the RGG experiment reconstructs only one graph G. All shortest path nodes $W_{\mathcal{P}^*_{ij}}$ in graph G are predicted as shortest path node candidates by our RGG strategy.

To evaluate the accuracy of our geometry-based method, we use the statistical precision metric [69,70]. Our evaluation procedure can be summarized as follows:

(1) Generate an SRGG G with required parameters N, $\mathbb{E}[D]$, and β. Select a number of node pairs in G uniformly at random.

(2) For each (i, j) node pair, we find all the shortest path nodes $W_{\mathcal{P}^*_{ij}}$. Nodes in $W_{\mathcal{P}^*_{ij}}$ are regarded as ground-truth set.

(3) Obtain the geodesic $\gamma(i, j)$ and compute the distance to the geodesic $d(q, \gamma(i, j))$ for all nodes, excluding nodes i and j.

(4) Determine the set of $\mathcal{S}$ nodes with the smallest distance-to-geodesic values, such that $|\mathcal{S}| = |W_{\mathcal{P}^*_{ij}}|$.

(5) Compute the precision $= |\mathcal{S} \bigcap W_{\mathcal{P}^*_{ij}}|/|\mathcal{S}|$.

We additionally analyze the area under the precision-recall curve (AUPR) as a supplementary performance metric to provide a more comprehensive evaluation of the method [71].

III. RESULTS

A. Quantifying the alignments of shortest paths along geodesic curves

We begin our work with the analysis of the alignment of shortest paths along geodesic curves in SRGGs. We generate an instance of an SRGG of $N = 10^4$ nodes, connected with Fermi-Dirac connection probabilities given by Eq. (5) with inverse temperature $\beta = 4$, and parameter d_0 corresponding to the expected degree of $\mathbb{E}[D] = 5$. Figure 2(b) depicts the distributions of the maximum, the minimum, and the average distance to geodesic from shortest path nodes for 10^6 randomly selected path end points within the SRGG, indicating that all three measures are substantially smaller than expected by chance. Our observation is not specific to the choice of SRGG parameters. We observe similar shortest path alignment properties in SRGGs with different parameters (Fig. 8), and the same pattern persists when different link connection probability functions are used (Fig. 11).

After establishing that shortest path nodes are closer to geodesics than expected by chance, we next investigate how topological properties of SRGGs affect path alignments. We measure the average distance to geodesic $\langle d \rangle$ and the average shortest path stretch $\langle S \rangle$, as described in Sec. II B. For each set of SRGG parameters, we generate 100 SRGG realizations if $N \leqslant 100$ and 1 realization for $N > 100$. For each generated graph G_i, $1 \leqslant i \leqslant 100$, we consider 1000 shortest paths between randomly selected connected node pairs when $N > 100$, and all connected node pairs in graphs with $N \leqslant 100$.

Figure 3 depicts the average shortest path distance to geodesic $\langle d \rangle$ and the average shortest path stretch $\langle S \rangle$ as a function of inverse temperature β and the average degree $\langle D \rangle$.

We observe that the average distance to the geodesic and the average path stretch decrease as β increases [Figs. 3(a) and 3(c)]. This observation is expected. Indeed, by design, larger inverse temperature β values favor short-distance connections in the SRGG [see Eq. (5)]. As a result, shortest path nodes are located closer to each other and closer to

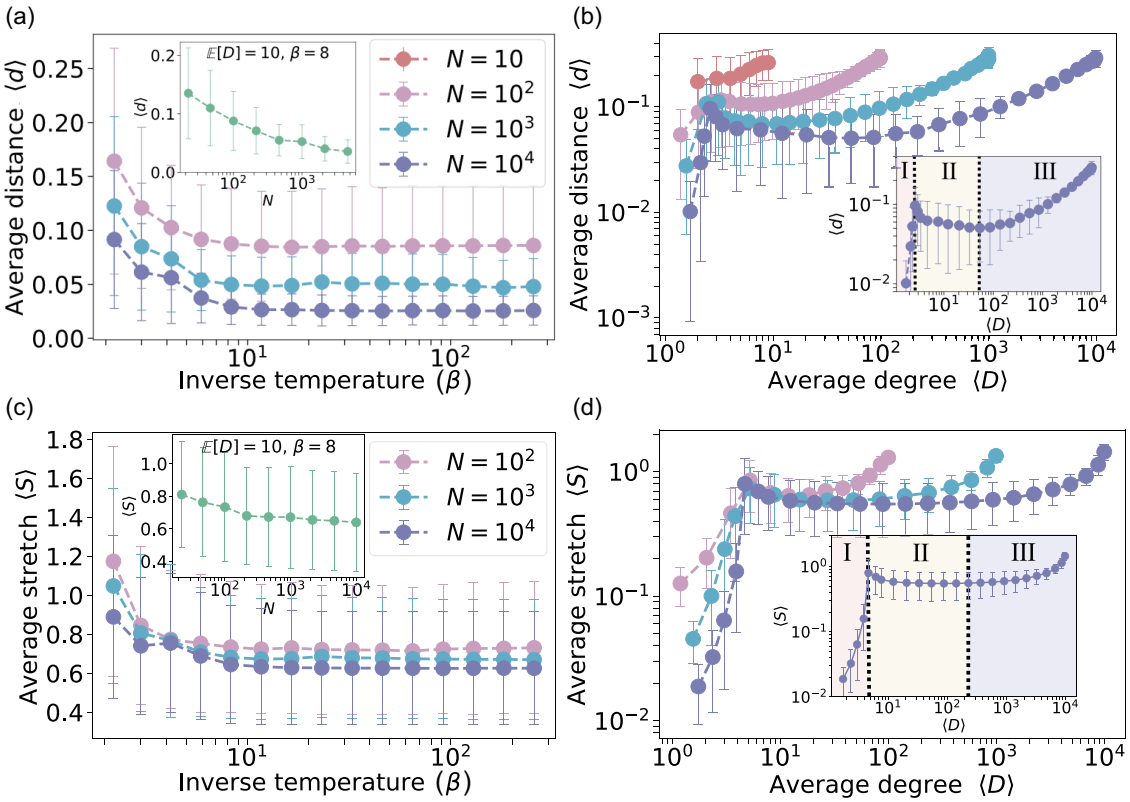


FIG. 3. (a) The average distance to the geodesic $\langle d \rangle$ as a function of inverse temperature β for SRGGs with $\mathbb{E}[D] = 10$ and variable size N. The inset displays the average distance to the geodesic $\langle d \rangle$ as a function of the number of nodes N for SRGGs with expected degree $\mathbb{E}[D] = 10$ and the inverse temperature $\beta = 8$. (b) The average distance to the geodesic $\langle d \rangle$ as a function of the average degree $\langle D \rangle$ for SRGGs of variable size N and the inverse temperature $\beta = 4$. The inset depicts the three phases discussed in the text. (c) The average shortest path stretch $\langle S \rangle$ as a function of the inverse temperature β for SRGGs of $\mathbb{E}[D] = 10$ and variable size N. The inset displays the average shortest path stretch $\langle S \rangle$ as a function of the number of nodes N for SRGGs with expected degree $\mathbb{E}[D] = 10$ and the inverse temperature $\beta = 8$. (d) The average shortest path stretch $\langle S \rangle$ as a function of the average degree $\langle D \rangle$ for SRGGs of variable size N and the inverse temperature parameter $\beta = 128$. All plots correspond to the average of 1000 randomly selected shortest paths, while the error bars quantify the standard deviations.

the geodesic connecting path end points, decreasing the average distance to the geodesic and the average shortest path stretch.

We also observe that the average distance to geodesic $\langle d \rangle$ and the average stretch $\langle S \rangle$ decrease as the network size N increases, if the remaining SRGG parameters are held constant. Indeed, the density of nodes in the latent space increases as the network size N increases. Therefore, distances between connected nodes must decrease on average to preserve the expected degree $\mathbb{E}[D]$. It is the decrease of distances between connected nodes that leads to the decrease of the average distance to geodesic $\langle d \rangle$ and the average shortest path stretch $\langle S \rangle$, observed in Figs. 3(a) and 3(c).

Additionally, the average distance to geodesics $\langle d \rangle$ and the average path stretch $\langle S \rangle$ exhibit a nonmonotonic behavior as a function of the graph average degree $\langle D \rangle$ as demonstrated in Figs. 3(b) and 3(d). Initially, $\langle d \rangle$ and $\langle S \rangle$ increase as the average degree $\langle D \rangle$ increases, Phase I in Figs. 3(b) and 3(d). After reaching the local maximum, $\langle d \rangle$ and $\langle S \rangle$ tend to decrease as the average degree increases, Phase II, reaching minimum values $\langle d_{\min} \rangle$ $\langle S_{\min} \rangle$. After reaching local minima, the average distance $\langle d \rangle$ and the average stretch $\langle S \rangle$ resume their growth as a function of average degree in Phase III, as seen in Figs. 3(b) and 3(d).

We hypothesize that the initial growth of the average distance $\langle d \rangle$ and the average shortest path stretch $\langle S \rangle$ in Phase I is related to the onset of the giant connected component (GCC) in the SRGGs. Indeed, when the average degree $\langle D \rangle$ is small, the network is fragmented into a large number of small disjoint subgraphs [Fig. 13(b)]. As the average degree increases, disjoint subgraphs eventually merge into a GCC of a size comparable to that of the entire graph. Right above the critical value of $\langle D \rangle = \langle D \rangle_c$ corresponding to the onset of the GCC, most of the nodes in the graph are connected, but the shortest paths between them may be arbitrarily long [see Fig. 13(a) and Appendix C]. As a result, the local maximum in the distance to geodesic $\langle d \rangle$ and the shortest path stretch $\langle S \rangle$ at the border of Phases I and II may coincide with the critical value $\langle D \rangle_c$ corresponding to the emergence of the GCC. To check if this is the case, we compared the $\langle D \rangle$ values corresponding to the local maximum of the distance to geodesic $\langle d \rangle$ with those corresponding to the GCC emergence for a rangle of SRGG parameters. As seen in Fig. 4(b), the two values are nearly equal, leading to a strong correlation, Pearson $R = 0.99$. The results support our hypothesis that the growth of the distance to geodesic $\langle d \rangle$ and the shortest path stretch $\langle S \rangle$ in Phase I are driven by the formation of the giant connected component in SRGG.

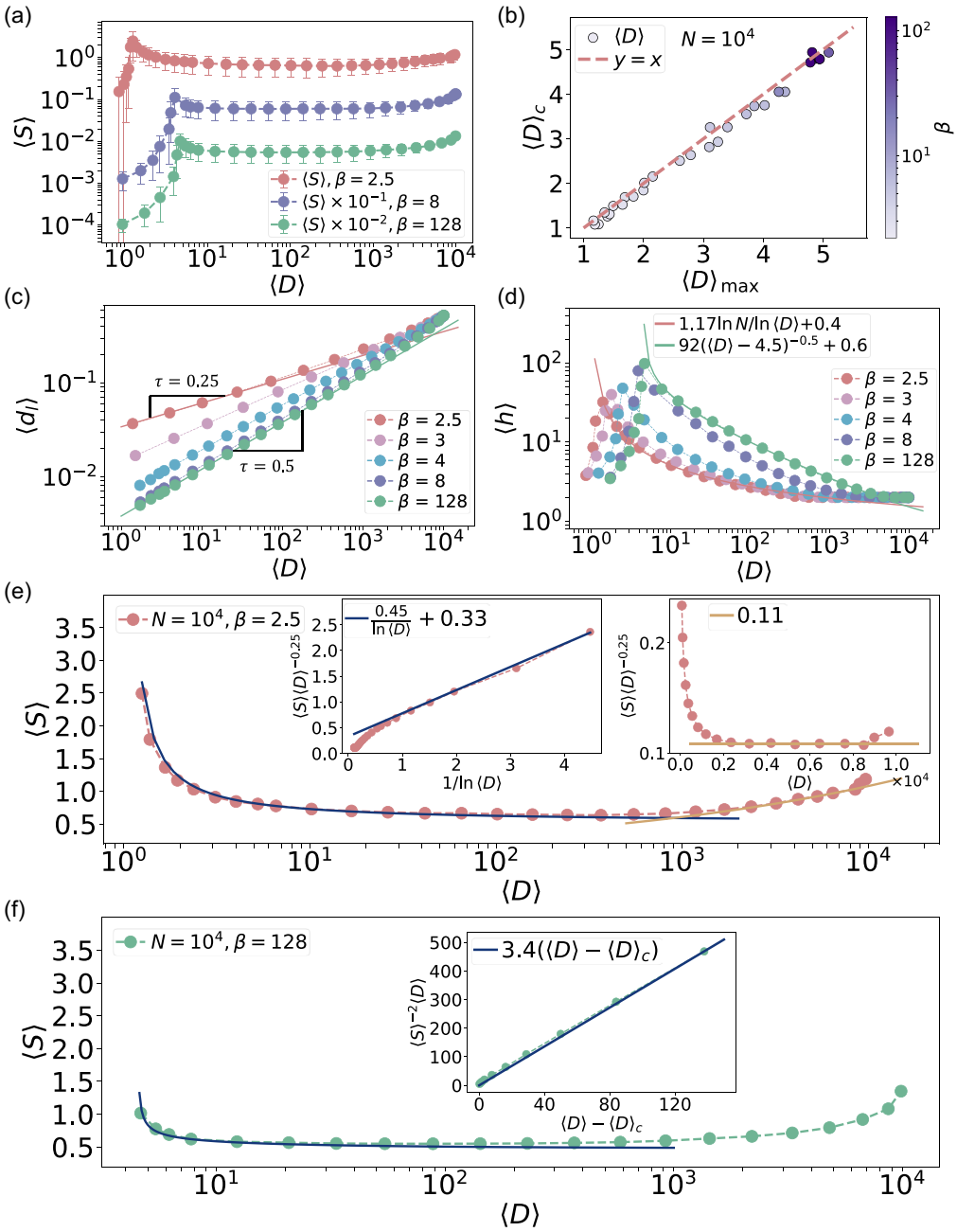


FIG. 4. (a) The average shortest path stretch $\langle S\rangle$ as a function of the average degree $\langle D\rangle$ for inverse temperatures $\beta = 1$, $\beta = 8$, and $\beta = 128$. The plots for beta $\beta = 8$ and $\beta = 128$ are scaled by multiplicative factors 10^{-1} and 10^{-2}, respectively, to avoid their overlap. The error bars depict standard deviation values. (b) Critical degree values $\langle D\rangle_c$ corresponding to the onset of the GCC as a function of expected degrees $\langle D\rangle_{\max}$ corresponding to the largest distance to geodesic. The scatter plot includes 20 points corresponding to different inverse temperature values β selected uniformly at random from the (2, 128] interval. (c) The average link length $\langle d\rangle_\ell$ as a function of the average degree $\langle D\rangle$ in SRGGs with different inverse temperatures β. The solid lines are power-law curves with exponents $\tau = \min(1/2, \beta/2 - 1)$. (d) The average shortest path hop count $\langle h\rangle$ as a function of the average degree $\langle D\rangle$ in SRGGs with different inverse temperatures β. The solid lines are empirical fits for $\beta = 2.5$ and $\beta = 128$ values. The average shortest path stretch $\langle S\rangle$ as a function of the average degree $\langle D\rangle$ for inverse temperatures (e) $\beta = 2.5$ and (f) $\beta = 128$. (e) The left inset depicts $\langle S\rangle\langle D\rangle^{-\tau}$ as a function of $\ln\langle D\rangle^{-1}$. Note that $\langle S\rangle\langle D\rangle^{-\tau} \propto \ln\langle D\rangle^{-1}$, confirming the low-degree behavior of the shortest path stretch $\langle S\rangle$. The right inset depicts $\langle S\rangle\langle D\rangle^{-\tau}$ as a function of $\langle D\rangle$. Note that $\langle S\rangle\langle D\rangle^{-\tau}$ reaches a horizontal asymptote, confirming the power-law scaling of the shortest path stretch for large $\langle D\rangle$ values. (f) The inset depicts $\langle S\rangle^{-2}\langle D\rangle$ as a function of $\langle D\rangle - \langle D\rangle_c$, testing the low-degree behavior of the shortest path stretch. Note that the shortest path stretch $\langle S\rangle$ is nearly independent of $\langle D\rangle$ for a large range of $\langle D\rangle$ values.

After reaching the local maximum, as seen in Figs. 3(b) and 3(d) the average path distance to geodesic $\langle d \rangle$ and the average path stretch $\langle S \rangle$ decrease as the average degree $\langle D \rangle$ increases until they reach their local minima at the border of Phases II and III.

We explain these nonmonotonic behaviors of $\langle S \rangle$ and $\langle d \rangle$ as resulting from the interplay of two competing effects. On the one hand, as the average degree increases, the average number of hops of a shortest path decreases, improving its alignment along a geodesic. On the other hand, larger average degrees result in larger distances between connected nodes, leading to larger deviations from the geodesic and larger stretch values. Based on Figs. 3(b) and 3(d), which show the nonmonotonic dependence as the network size grows, we expect this behavior to persist in the asymptotic limit, i.e., as $N \to \infty$.

Although an exact computation of the shortest path stretch is practically infeasible due to multiple sources of correlation, we can still quantify the previous effects using a mean-field-type approximation. Within this approach, the stretch is simply expressed as the product of the average link length $\langle d_l \rangle$ and the average hop count $\langle h \rangle$,

$$\langle S \rangle \approx \langle d_l \rangle \langle h \rangle. \tag{10}$$

More details about the above approximations can be found in Appendix F.

We estimate the average link length $\langle d_l \rangle$ using Bayes' rule. The conditional probability $\rho(D_{ij} = x | A_{ij} = 1)$ that the distance D_{ij} between two randomly chosen nodes i and j equals x, given that the two nodes are connected, $A_{ij} = 1$, depends on the conditional probability nodes i and j are connected, given that the distance between them equals x, $P(A_{ij} = 1 | D_{ij} = x)$, the probability density $\rho(D_{ij} = x)$ for a distance between two randomly chosen nodes in the unit square, and the probability $P(A_{ij} = 1)$ that two randomly chosen SRGG nodes are connected. Since $P(A_{ij} = 1 | D_{ij} = x)$ is nothing else but the connection probability function $f(x)$ prescribed by Eq. (5), we have

$$\rho(D_{ij} = x | A_{ij} = 1) = \frac{f(x)\rho(D_{ij} = x)}{P(A_{ij} = 1)}. \tag{11}$$

Then, the average link length $\langle d_l \rangle$ can be estimated (see Appendix D) as

$$\langle d_l \rangle = \frac{\int x f(x) \rho(D_{ij} = x) \mathrm{d}x}{\int f(x) \rho(D_{ij} = x) \mathrm{d}x} \sim \langle D \rangle^{\tau}, \tag{12}$$

where $\tau = 1/2$ for $\beta > 3$ and $\tau = \beta/2 - 1$ for $\beta \in (2, 3)$. For the numerical verification of the derived scaling relation appearing in Eq. (12), see Fig. 4(c).

Figure 4(d) depicts the behavior of the average hop count $\langle h \rangle$ as a function of the average degree $\langle D \rangle$. For large inverse temperatures β, SRGG is similar to an RGG, and we can approximate the behavior of the average hop count as

$$\langle h \rangle \approx \frac{\langle \delta \rangle}{[\langle D \rangle - \langle D \rangle_c]^{1/2}} + C, \quad \beta \gg 3, \tag{13}$$

where $\langle \delta \rangle$ is the average distance between two randomly chosen points in the unit square, $\langle D \rangle_c \approx 4.51$ is the critical value of the average degree corresponding to the onset of the GCC [50], and C is a constant. For small inverse temperatures β [see Fig. 4(d)], SRGG is akin to a configurational model, and

$$\langle h \rangle \approx \frac{\ln N}{\ln \langle D \rangle} + A, \quad \beta \in (2, 3), \tag{14}$$

where N is the number of nodes and for networks with Binomial degree distribution $A = \frac{1}{2} - \frac{\gamma - 2\ln(1 - 1/\langle D \rangle)}{\ln \langle D \rangle} - \frac{2\,\mathbb{E}[\ln W | W > 0]}{\ln \langle D \rangle}$. Here, γ is the Euler constant and W is the almost sure limit of the normalized branching process [72,73]. For graphs with expected degree $\mathbb{E}[D] > 1.3$ and a binomial degree distribution, $\mathbb{E}[\ln W \mid W > 0] \in (-0.2, 0.2)$ [72]. Since $\mathbb{E}[\ln W \mid W > 0]$ is difficult to measure numerically, in this work we estimate the entire constant A by fitting $\langle h \rangle$ versus $\langle D \rangle$ curves [see Fig. 4(d)].

The scaling behaviors of $\langle d_l \rangle$ and $\langle h \rangle$ given in Eqs. (12)–(14) account for the observed dependence of the average stretch $\langle S \rangle$ on the average degree $\langle D \rangle$ in Phases II and III.

For small inverse temperatures, $2 < \beta < 3$, the average stretch scales as $\langle S \rangle \propto (\frac{\ln N}{\ln \langle D \rangle} + C) \langle D \rangle^{\frac{\beta}{2} - 1}$. Consequently, at small $\langle D \rangle$ (the onset of Phase II), the stretch exhibits the asymptotic behavior $\langle S \rangle \sim \langle D \rangle^{\beta/2 - 1} / \ln \langle D \rangle$ as seen in the left inset of Fig. 4(e), whereas at large $\langle D \rangle$ (deep in Phase III), the logarithmic correction becomes negligible and $\langle S \rangle \sim \langle D \rangle^{\beta/2 - 1}$ [see the right inset of Fig. 4(e)].

For large inverse temperatures, $\beta \gg 3$, the average stretch follows $\langle S \rangle \approx (\frac{\langle \delta \rangle}{[\langle D \rangle - \langle D \rangle_c]^{1/2}} + C) \langle D \rangle^{1/2}$. As a consequence, for $\langle D \rangle \gg \langle D \rangle_c$ the dependence on $\langle D \rangle$ is weak, yielding stretch values that are approximately constant, as seen in Fig. 4(f).

Overall, our results indicate that shortest paths in soft random geometric graphs are aligned along geodesic curves connecting shortest path end points. The alignment, as quantified by the average distance to geodesic $\langle d \rangle$ and the shortest path stretch $\langle S \rangle$, increases as the inverse temperature β and the SRGG size N increase. In contrast, the dependence of shortest path alignment on the average degree $\langle D \rangle$ is nonmonotonic, reflecting the competition between two opposing effects: the reduction of the average hop count $\langle h \rangle$ and the concurrent increase of the average link length $\langle d_\ell \rangle$ as connectivity increases.

B. Finding shortest path nodes in soft random geometric graphs

After observing and quantifying the alignment of shortest paths along geodesic curves in soft random geometric graphs, we next ask if shortest path nodes are identifiable based on their proximity to geodesic curves. To identify shortest path nodes connecting an $i - j$ node pair of interest, we draw a geodesic line $\gamma(i, j)$ connecting path end points i and j and then for every other network node $\ell \neq i \neq j$ we compute its distance to geodesic $d(\ell, \gamma(i, j))$. The smaller the distance, the higher the likelihood for node ℓ to be a shortest path node (see Sec. II).

Figure 5 displays the results of our path-finding experiments for a combination of inverse temperature β and expected degree $\mathbb{E}[D]$ parameters. We observe that precision score for the distance to the geodesic in finding shortest path nodes is significantly larger than expected by chance. Figure 14 further illustrates the AUPR as a supplementary evaluation of the geometry-based method, showing results consistent with the precision scores.

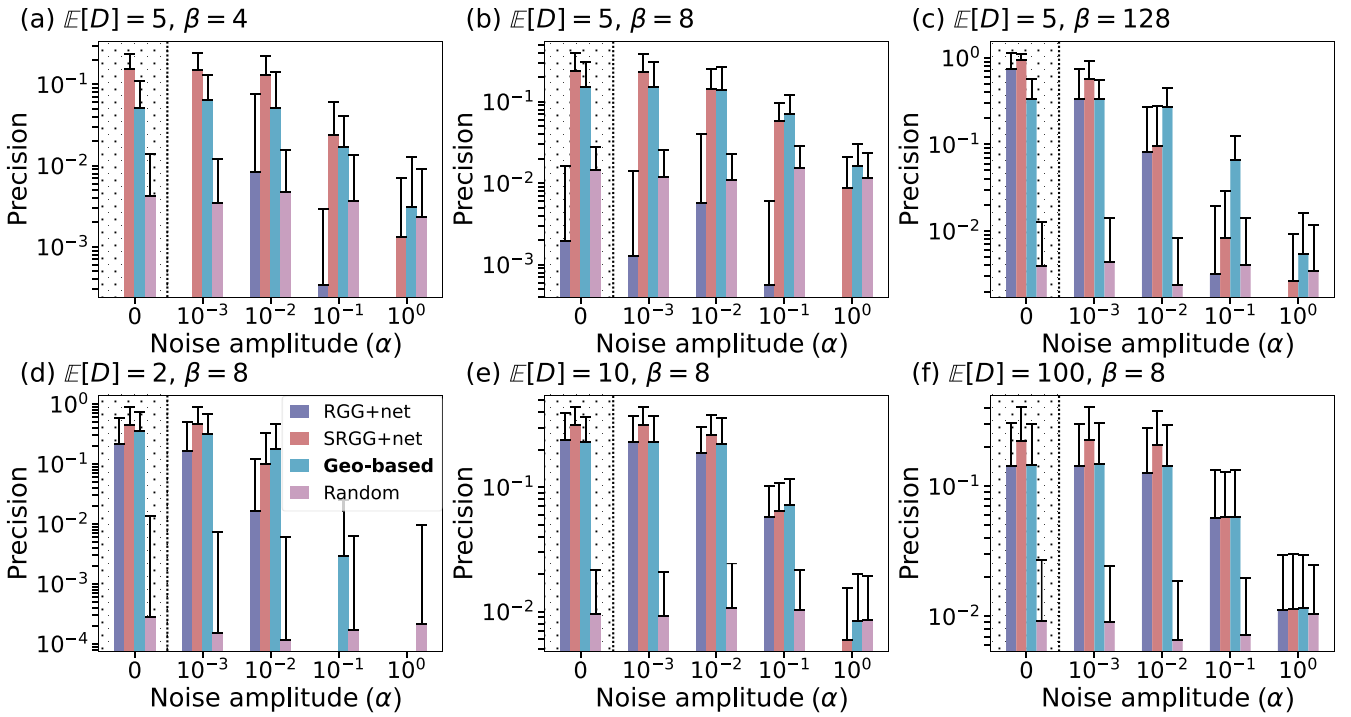


FIG. 5. Precision of shortest path finding using (1) distance to geodesic (geo-based), (2) SRGG (SRGG+network), and (3) RGG (RGG+network) reconstructions. Panels display average precision as a function of the coordinate noise amplitude α and panels correspond to different combinations of $\mathbb{E}[D]$ and β parameters. All experiments are performed using SRGG networks of $N = 10^4$ nodes. Each bar is the average precision computed for 100 randomly chosen node pairs and the error bars depict standard deviations.

Inspired by this observation, we ask if there are plausible circumstances in which the distance to the geodesic can find shortest path nodes with accuracy higher than known state-of-the-art methods. Clearly, if network G is fully observable, existing graph theoretic algorithms, e.g., the Dijkstra algorithm, can find shortest path nodes precisely and in polynomial time. A much more challenging situation is when only network nodes and their approximate positions are known in network G. In more precise terms, we assumed that SRGG links are not known. At the same time, all SRGG nodes are known, but their coordinates are blurred with random noise of amplitude $\alpha > 0$:

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \alpha X_i, \quad \tilde{\mathbf{y}}_i = \mathbf{y}_i + \alpha Y_i, \tag{15}$$

where $i = 1, \ldots, N$, and X_i and Y_i are random variables drawn from the uniform distribution $U[-1, 1]$.

In this situation, one can still find shortest path nodes using the distance to the geodesic method with blurred node coordinates $\{\tilde{\mathbf{x}}_i, \tilde{\mathbf{y}}_i\}$. To find shortest path nodes with the Dijkstra network-based algorithm, one needs to first reconstruct SRGG links. To this end, we consider two strategies. The SRGG+net strategy presumes that we know the connection probability function $f(x)$ used to generate the original SRGG graph. In this case, one first generates an SRGG instance with the same connection probability function $f(x)$ and then applies the Dijkstra algorithm to the resulting graph to find shortest path nodes. Another strategy, which we call RGG+net, assumes that the average degree $\langle D \rangle$ of the graph is known, but the connection probability function $f(x)$ is not. In this case, we treat the graph at hand as a random geometric graph, we compute its effective connectivity radius R based on the number of nodes and the average degree, and then connect all nodes if their blurred positions are closer than R (see Sec. II C). In Table I, we summarize and compare the information utilized by each method. Specifically, the SRGG+net baseline is given full knowledge of the connection function, including the network size, node coordinates, inverse temperature, and expected degree, whereas the RGG+net baseline assumes less model specification, lacking knowledge of the inverse temperature. The geometry-based method, in contrast, does not reconstruct links at all and uses only proximity to the geodesic.

By comparing the precision of the distance to geodesics with those of the alternative SRGG+net and RGG+net strategies, we make several observations. When node coordinates are known precisely, $\alpha = 0$, the SRGG+net and RGG+net strategies offer substantially higher precision scores than those of the distance to the geodesic, as shown in Fig. 5. This result is expected since precisely known node coordinates

TABLE I. Information utilized by different shortest path node-finding methods.

Method	Network size	Node coordinates	Expected degree	Inverse temperature
RGG+net	✓	✓	✓	×
SRGG+net	✓	✓	✓	✓
Geo-based	✓	✓	×	×
Random	✓	×	×	×

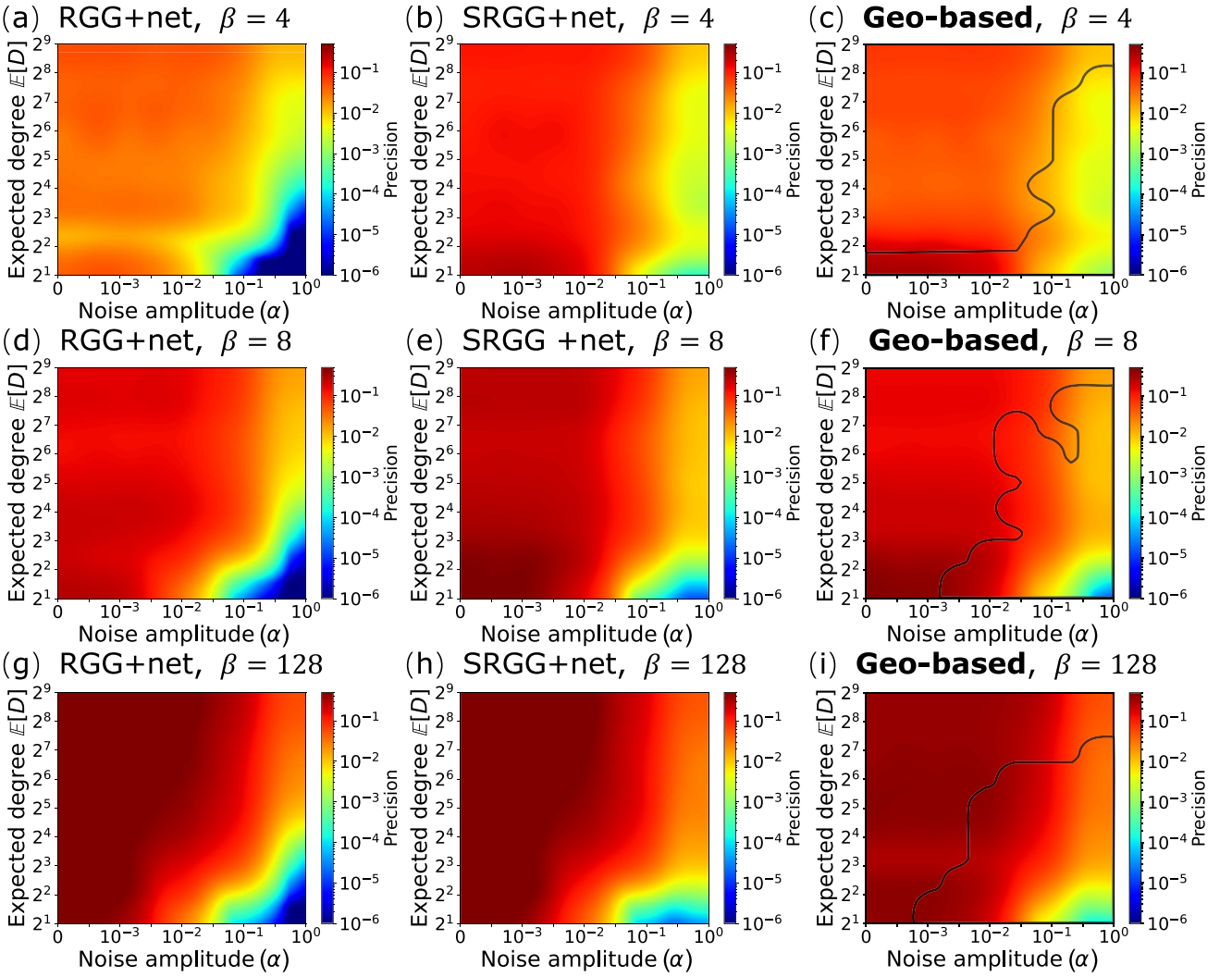


FIG. 6. Heatmap of precision. The x axis is the noise amplitude, while the y axis represents the expected degree. For each expected degree $\mathbb{E}[D]$ and inverse temperature β, a 10^4-node SRGG is generated and the coordinates are, respectively, blurred with noise amplitude α. The heatmaps consist of 9×9 points, each point corresponding to the average of 100 randomly selected node pairs in a (blurred) SRGG. In the region enclosed by the black line, the distance-to-geodesic approach (geo-based) achieves higher precision compared with the SRGG (SRGG+network) and RGG (RGG+network) reconstruction approaches.

allow for the reconstruction of graphs with relatively high accuracy.

Second, both SRGG+net and RGG+net strategies become more accurate in finding shortest path nodes as the inverse temperature β increases [Figs. 5(a)–5(c)]. We explain this observation by the decrease in randomness in SRGG networks: At higher β values, connections are preferentially established at shorter distances, and distinct realizations of reconstructed networks are less different. As a result, reconstructed networks G_ℓ better recreate the original network G, allowing for the identification of original shortest path nodes with higher precision, as shown in Fig. 5.

Third, the RGG+net strategy tends to be less accurate than the SRGG+net. However, the relative accuracy of the former seems to increase as the inverse temperature β increases. This result is also expected since larger β values favor shorter links, effectively making SRGGs closer to RGGs.

When comparing the distance to the geodesic approach to the SRGG+net and RGG+net strategies, we find that the precision of the distance-to-geodesic approach may exceed that of the SRGG+net and RGG+net when the noise amplitude α is large (Fig. 6), indicating that the distance-to-geodesic approach is more robust to coordinate uncertainties than network reconstruction approaches SRGG+net and RGG+net. In the case of noise magnitude α, distances between nodes may be perturbed by at most 2α and lead to changes in some node rankings with respect to distances to geodesics. The effects of node coordinates on network reconstruction approaches are potentially more disrupting. This is the case since link probabilities in the SRGG depend nonlinearly on distances between the nodes and even small changes in network links may significantly alter shortest paths. As a result, shortest paths in reconstructed networks may significantly differ from those in the original network, reducing the accuracy of network reconstruction strategies.

To explore the limits of shortest path finding with the distance to geodesic, we next conduct systematic experiments for ranges of noise amplitude $\alpha \in [0, 1]$, average degree $\mathbb{E}[D] \in [2, 2^9]$, and different inverse temperature values. As seen in Fig. 6, the accuracy of all path-finding methods decreases as the average degree $\mathbb{E}[D]$ decreases and coordinate uncertainty α increases. At the same time, the accuracy of the distance to the geodesic approach is less sensitive to noise, resulting in low average degree and high noise regions [Figs. 6(c), 6(f), and 6(i)], where the distance to the geodesic method is more accurate than network reconstruction strategies.

IV. DISCUSSION

In summary, we analyzed the alignment of shortest paths along geodesic lines connecting shortest path end points in SRGGs. Some of our results are intuitive and expected. In particular, we found that the alignments of shortest paths, as measured by the distance to the geodesic, and the path stretch become stronger in SRGGs as the number of nodes N and the inverse temperature β increase.

Nevertheless, the nonmonotonic behavior of shortest path alignment as a function of the average degree is far from trivial. As the average degree increases, two competing phenomena take place. On the one hand, the hop count of shortest paths decreases, limiting the extent of *wandering* of shortest paths. On the other hand, larger average degree values can only be achieved by allowing nodes to connect over larger *Euclidean* distances. As a result, distances between adjacent nodes in a shortest path increase, allowing for large distances from shortest path nodes to geodesic curves. The competition between these two effects results in the nonmonotonic behavior of the distance to the geodesic as a function of the average degree.

We found that the alignments of shortest paths along geodesic curves in *Euclidean* SRGGs are sufficiently strong to enable shortest pathfinding, even near criticality where deviation measures such as $\langle S \rangle$ and $\langle d \rangle$ reach their maximum values. Shortest path nodes in *Euclidean* SRGGs can be identified using distance to geodesic, similar to RHG [16]. The accuracy of such geometric shortest path finding is higher for graphs with a larger inverse temperature β, in agreement with our path alignment experiments.

The distance to the geodesic method uses node coordinates to find nodes closest to the geodesic connecting shortest path end points, and can be used in situations when node links are not known. In contrast, to enable traditional network-based methods in this situation, one needs first to reconstruct network links, which can be a highly nontrivial task. We find that the best advantage of the distance to the geodesic method is achieved for small average degree values and larger node coordinate uncertainties. While the geometric alignment of shortest path nodes is not necessarily the strongest in this case, network reconstruction methods are even less accurate, offering better prospects for using network geometry in finding shortest path nodes.

Our results are not specific to the choice of the connection probability function. While most of our analysis is based on SRGGs with the Fermi-Dirac connection probability function [Eq. (5)], we obtain qualitatively similar results for the alignment of shortest paths for Waxman SRGGs characterized by the exponentially decaying connection probabilities [Eq. (4) and see Fig. 11].

Our findings invite follow-up questions. One of them, inspired by real networks, is the problem of shortest path finding in geometric networks where nodes are distributed heterogeneously. Our preliminary results demonstrate that the shortest path nodes remain closer to the geodesic than expected by chance, even in SRGGs where nodes are *not* uniformly distributed (Fig. 12). While the distance to the geodesic can also be used in geometric networks with heterogeneous spatial node distributions, it is not clear if the accuracy of path finding remains high. We leave this question for future investigation.

Another open question outside the scope is the geometric properties of shortest paths in weighted geometric networks. While link weights can be functions of link distances, more elaborate models were proposed in the literature that decouple link weights from distances between connected nodes [74–76].

Finally, the important question is whether all shortest path nodes are equally difficult to find using distance to the geodesic. Our preliminary experiments indicate that shortest path nodes situated closer to path end points tend to have a smaller distance to the geodesic and, as a result, are easier

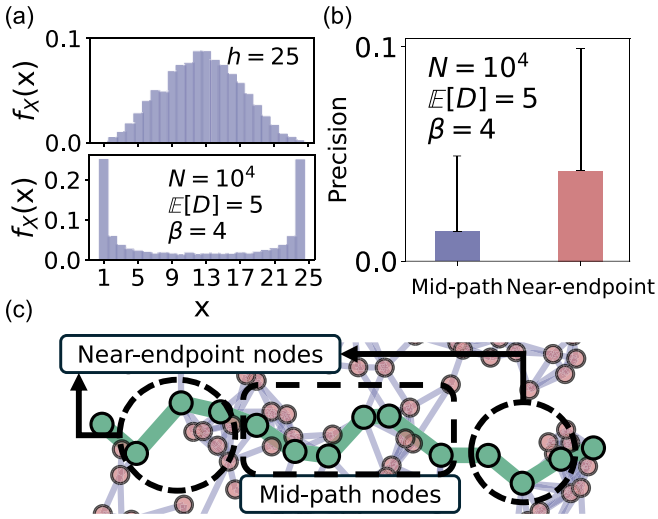


FIG. 7. [(a), top] The distribution of shortest path node positions with the largest distance to geodesic. [(a), bottom] The distribution of shortest path node positions with the smallest distance to geodesic. In both distributions, we consider shortest paths of fixed hop count, $h = 25$, and SRGGs with $N = 10^4$ nodes, inverse temperature $\beta = 4$, and the expected degree $\mathbb{E}[D] = 5$. The most remote from the geodesic nodes are likely to be found in the middle of a shortest path, while the closest to the geodesic nodes are likely to be found close to the path end points. See Appendix B for plots with other SRGG parameters. (b) The precision of finding shortest path nodes near end points (near-end-point) and in the middle of the path (mid-path) using the distance to the geodesic. A 10^4-node SRGG is generated. Each bar is the average precision computed for 100 randomly selected node pairs and the error bars depict standard deviations. Panel (c) shows an example illustrating the division of near-end-point nodes and mid-path nodes on a shortest path.

to infer (Fig. 7). This observation is exciting since shortest path nodes closer to the path end points may be easier to manipulate in networks compared to nodes in the middle of the shortest path. Indeed, nodes in the middle of a shortest path may be shared by other paths and can be involved in many vital processes. As a result, altering their connections or capacity may lead to cascades of changes throughout the network [77–80].

Figure 7 suggests that shortest path nodes closer to path end points are easier to find with the distance to the geodesic. Are nodes in the middle of a shortest path easier to find with network-based methods? This can indeed be the case since (1) nodes in the middle of a shortest path tend to have degrees larger than those close to its end points and (2) large degree nodes are often characterized by large betweenness centrality values [81,82]. As a result, nodes in the middle of shortest paths can be identified by prioritizing large-degree nodes. Thus, a fusion approach combining elements of the distance to geodesic with network topology may ultimately be more accurate in finding shortest path nodes than one of the two approaches used separately.

ACKNOWLEDGMENTS

M.K. has been supported by the Dutch Research Council (NWO) Grants No. OCENW.M20.244 and No. VI.C.242.106. P.V.M. was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 101019718). Z.Q. has been funded by the China Scholarship Council (Grant No. 202106070032)

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX A: NOTATION

Table II summarizes the principal notation that we use throughout this paper.

TABLE II. Notation.

Symbol	Definition
G	Graph or network
N	Number of nodes
L	Number of links
$l = i \sim j$	A link between nodes i and j
C_G	The average clustering coefficient of graph G
$\mathbb{E}[D]$	The expected degree of a graph
$\langle D \rangle$	The average degree of a graph
$\mathcal{P}_{ij}$	A path from node i to node j
$\mathcal{P}^*_{ij}$	A shortest path between nodes i and j
$W_{\mathcal{P}^*_{ij}}$	Nodes contained in all shortest paths connecting nodes i and j
h	The hop count
$\gamma(i, j)$	The geodesic curve between nodes i and j
d_{ij}	The geometric distance between nodes i and j
d_l	The geometric length of link l, i.e., the distance between its two adjacent nodes
$d(q, \gamma(i, j))$	The geometric distance from node q to the geodesic $\gamma(i, j)$
$\langle d \rangle$	The average distance to the geodesic
S	The shortest path stretch, i.e., the sum of its link lengths

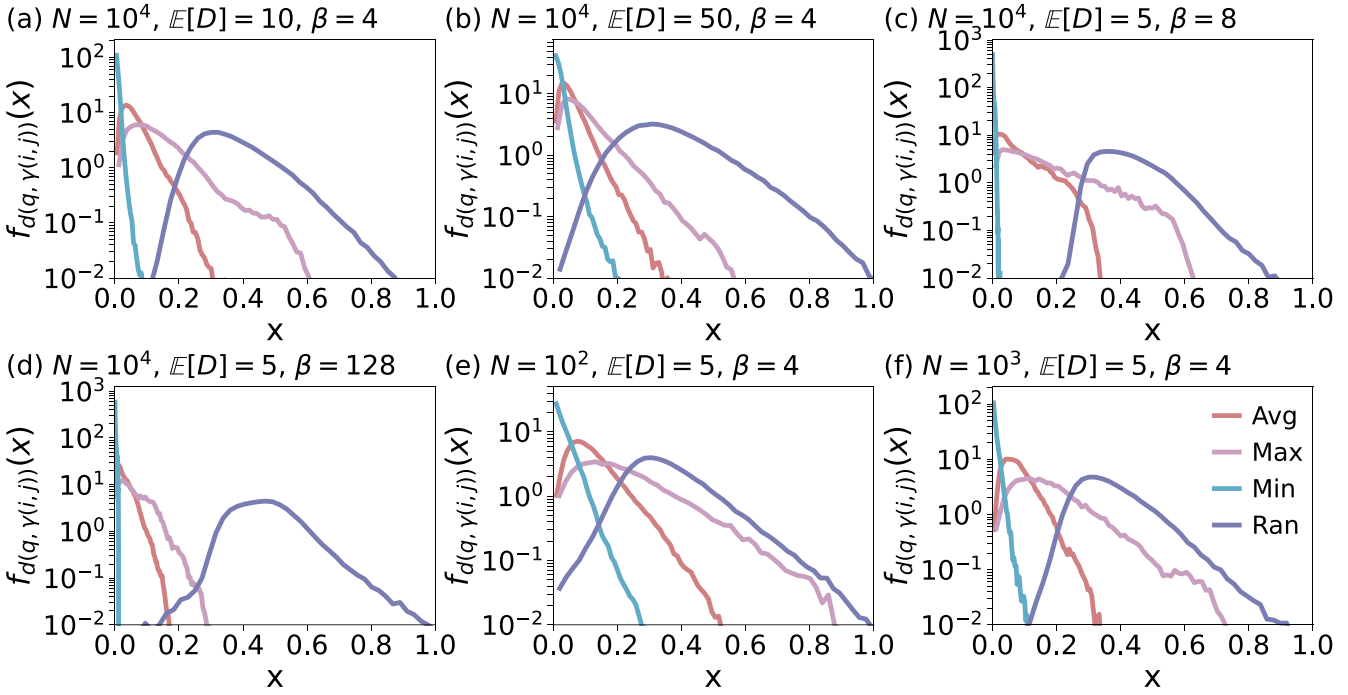


FIG. 8. The distribution of the average (avg), maximum (max), and minimum (min) distances between shortest path nodes and geodesics connecting shortest path end points. For comparison, we also plot the distribution of distances to geodesic from randomly selected nodes (ran). The distributions in panels (a)–(d) are obtained for a single SRGG realization and 10^6 random shortest paths. For panels (e) and (f), we generated 10^3 SRGG realizations. For each G_i, $i = 1, \ldots, 10^3$, we considered 10^3 random shortest paths.

APPENDIX B: DISTANCE TO GEODESIC

Figure 8 depicts the distributions of the distance to the geodesic in SRGGs for a variety of expected degree $\mathbb{E}[D]$ and inverse temperature β values. Figure 9 demonstrates the distribution of the node positions X in the shortest path corresponding to the largest distance to the geodesic. Figure 10 demonstrates the distribution of the node positions X in the shortest path corresponding to the minimum distance to the geodesic.

In this Appendix, we also study the distributions of the distance to the geodesic for different SRGG configurations, including uniform SRGGs with a Waxman connection probability function and SRGGs with nonuniformly distributed node coordinates and a Fermi-Dirac connection probability. First, we consider the Waxman connection probability, which is given by

$$f(d) = \beta e^{\left(-\frac{d}{d_0}\right)^{\eta}}. \tag{B1}$$

Waxman SRGGs of N nodes built on the unit square with uniform node density are characterized by the expected degree

$$\mathbb{E}[D] \approx 2\pi N\beta \frac{d_0^2}{\eta}\Gamma\left(\frac{2}{\eta}\right). \tag{B2}$$

Figure 11 depicts the distributions of the distance to the geodesic for a variety of expected degree $\mathbb{E}[D]$ and parameter η values. Note that the results for the Waxman SRGGs are similar to SRGGs with the Fermi-Dirac connection probability function.

Figure 12 further presents the distributions of the distance to the geodesic in SRGGs where node coordinates are not uniformly scattered. Specifically, we sample node positions from a Gaussian mixture with randomly placed cluster centers and heterogeneous variances, introducing controlled spatial inhomogeneity into the SRGG. We first generate the coordinates of ten cluster centers, where each center (x_k, y_k), for $k = 1, \ldots, 10$, is independently and uniformly sampled from $(0, 1)$. Each cluster is then assigned a standard deviation σ_k. For each node i, a cluster k is randomly selected and the node coordinates x_i and y_i are, respectively, sampled from Gaussian distributions centered at x_k and y_k with standard deviation σ_k. Finally, given the generated node coordinates, links in the SRGG are determined according to the connection probability defined in Eq. (12). The results shown in Fig. 12 remain consistent with those observed in SRGGs where nodes are uniformly distributed (Fig. 8).

APPENDIX C: THE ONSET OF THE GIANT CONNECTED COMPONENT IN THE SRGG

In Fig. 13, we show the onset of the GCC in the SRGGs.

APPENDIX D: THE AVERAGE LINK LENGTH

To derive the expression for the average link length $\langle d_l \rangle$ in SRGG models, we first define $\rho(D_{ij} = x | A_{ij} = 1)$ as the conditional probability density that the distance D_{ij} between two randomly chosen nodes in the SRGG equals x, provided

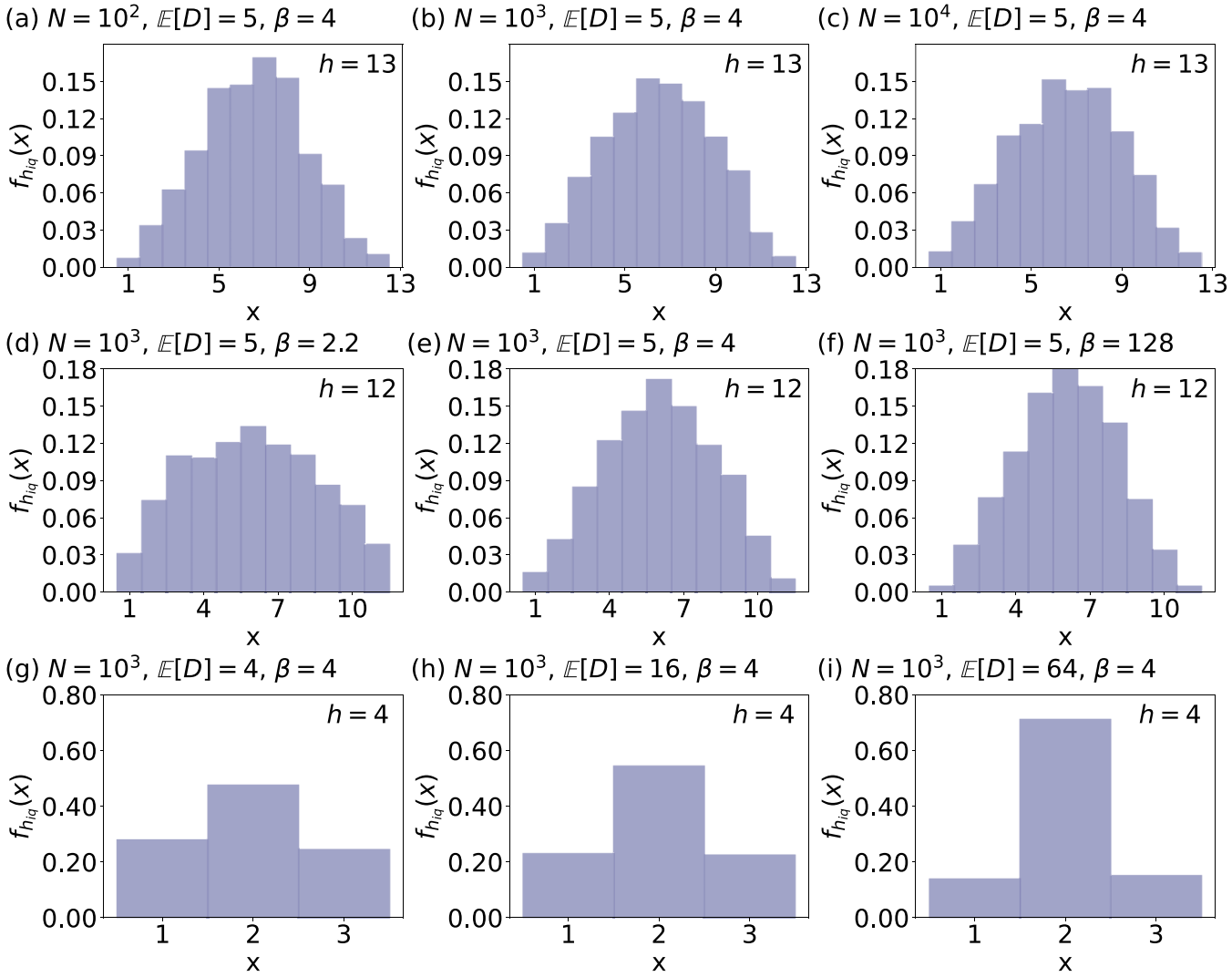


FIG. 9. Distribution of the position X of the node with the largest distance to the geodesic along shortest paths in SRGGs for different combinations of network size N, inverse temperature β, and the expected degree $\mathbb{E}[D]$. In each panel, only shortest paths with a fixed hop count h are considered. For each shortest path, the node with the maximum distance to the geodesic is identified, and its position $X \in \{1, \ldots, h\}$ along the path is recorded. Panel (a) averages over 10^3 independent SRGG realizations, while the remaining panels correspond to a single SRGG realization. In all cases, statistics are collected from 10^5 randomly selected shortest paths per SRGG realization.

these nodes are connected, $A_{ij} = 1$. The average link length is then given by $\langle d_l \rangle = \int_0^{\sqrt{2}} x\rho(D_{ij} = x|A_{ij} = 1)\mathrm{d}x$.

Using the Bayes' rule, the thought pdf can be expressed as

$$\rho(D_{ij} = x|A_{ij} = 1) = \frac{P(A_{ij} = 1|D_{ij} = x)\rho(D_{ij} = x)}{P(A_{ij} = 1)}, \quad \text{(D1)}$$

where $P(A_{ij} = 1|D_{ij} = x)$ is nothing else but the connection probability function $f(x)$, prescribed by Eq. (5), and $\rho(D_{ij} = x)$ is the distribution of distances between node pairs in a unit square. $P(A_{ij} = 1)$ is the probability that a randomly chosen node pair is connected, $P(A_{ij} = 1) = \frac{2L}{N(N-1)} = \frac{\langle D \rangle}{N-1}$.

The average link length is then given by

$$\langle d_l \rangle = \frac{\int_0^{\sqrt{2}} xf(x)\rho(D_{ij} = x)\mathrm{d}x}{\int_0^{\sqrt{2}} f(x)\rho(D_{ij} = x)\mathrm{d}x}. \quad \text{(D2)}$$

We approximate the distribution of distances between node pairs in a unit square as $\rho(D_{ij} = x) = \frac{2x}{R^2}$, where R is the size of a unit square, $R = \mathcal{O}(1)$, obtaining

$$\langle d_l \rangle \approx d_0 \frac{\int_0^{(\sqrt{2}/d_0)^\beta} \frac{u^{3/\beta-1}}{1+u}\mathrm{d}u}{\int_0^{(\sqrt{2}/d_0)^\beta} \frac{u^{2/\beta-1}}{1+u}\mathrm{d}u}. \quad \text{(D3)}$$

The closed-form expression for $\langle d_l \rangle$ in Eq. (D3) is given by the ratio of two hypergeometric functions. It is useful, however, to infer its leading-order behavior as a function of $\langle D \rangle$. To this end, we recall that $\frac{1}{d_0} \sim N^{1/2} \gg 1$ for large N. This follows from Eq. (6) for sparse SRGGs.

Since $\beta > 2$, the integral in the denominator of Eq. (D3) can be approximated as $\int_0^\infty \frac{u^{2/\beta-1}}{1+u}\mathrm{d}u = \pi/\sin(\frac{2\pi}{\beta})$. Further, if $\beta > 3$, the integral in the numerator of Eq. (D3) can be approximated in a similar manner as $\int_0^\infty \frac{u^{3/\beta-1}}{1+u}\mathrm{d}u = \pi/\sin(\frac{3\pi}{\beta})$.

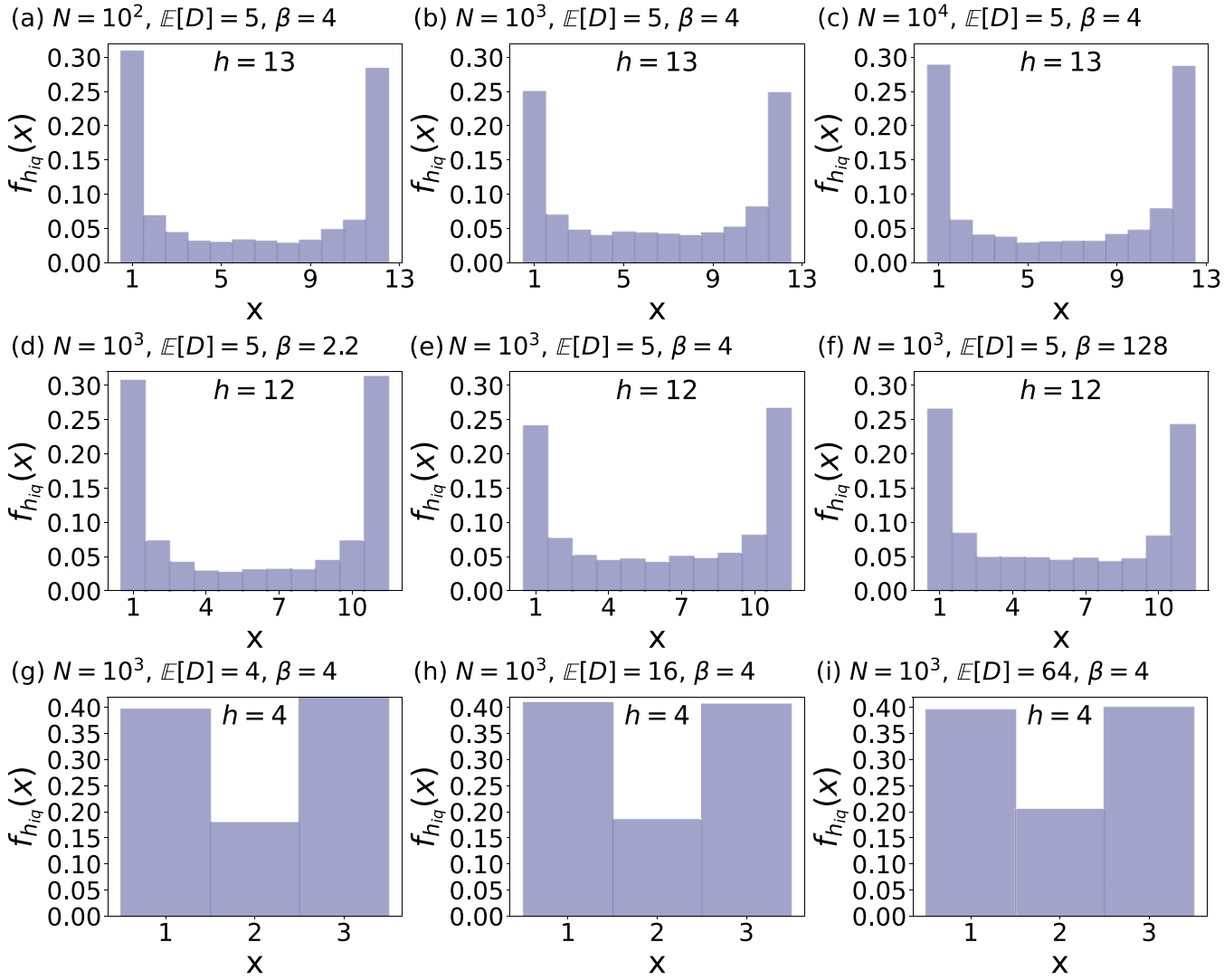


FIG. 10. Distribution of the position X of the node with the smallest distance to the geodesic along shortest paths in SRGGs for different combinations of network size N, inverse temperature β, and the expected degree $\mathbb{E}[D]$. In each panel, only shortest paths with a fixed hop count h are considered. For each shortest path, the node with the minimum distance to the geodesic is identified, and its position $X \in \{1, \ldots, h\}$ along the path is recorded. Panel (a) averages over 10^3 independent SRGG realizations, while the remaining panels correspond to a single SRGG realization. In all cases, statistics are collected from 10^5 randomly selected shortest paths per SRGG realization.

As a result,

$$\langle d_l \rangle \approx d_0 \frac{\sin\left(\frac{2\pi}{\beta}\right)}{\sin\left(\frac{3\pi}{\beta}\right)} \sim \langle D \rangle^{1/2}, \quad \beta > 3. \tag{D4}$$

In case $\beta \in (2, 3)$, we approximate the integral in the numerator of Eq. (D3) as $\int_0^{(\sqrt{2}/d_0)^\beta} u^{3/\beta-2} \mathrm{d}u$, which results in

$$\langle d_l \rangle \approx \frac{\pi \beta 2^{\frac{3-\beta}{2}} \sin\left(\frac{2\pi}{\beta}\right)}{3-\beta} d_0^{\beta-2} \sim \langle D \rangle^{\beta/2-1}, \quad 2 < \beta < 3. \tag{D5}$$

APPENDIX E: EVALUATION OF GEOMETRIC BASED METHOD USING AUPR

Since the geometry-based method produces an ordering of nodes according to their distance to the geodesic, we further assess its performance using the AUPR. In Fig. 14, we report the results of our path-finding experiments for different combinations of the inverse temperature β and the expected degree $\mathbb{E}[D]$ in SRGGs, where node coordinates are perturbed by random noise [Eq. (15)]. Specifically, the experiments are executed as follows:

(1) Generate an SRGG G with given parameters N, $\mathbb{E}[D]$, and β. Node coordinates are perturbed by random noise according to Eq. (15). A set of node pairs is then selected uniformly at random.

(2) For each node pair (i, j), we determine the set of nodes lying on shortest paths, $W_{\mathcal{P}_{ij}^*}$, which is used as the ground-truth set.

(3) Compute the geodesic $\gamma(i, j)$ and evaluate the distance $d(q, \gamma(i, j))$ for all nodes $q \in V \setminus \{i, j\}$.

(4) Compute the AUPR using the distance-to-geodesic ranking and the corresponding ground-truth shortest path node sets.

The SRGG+net method (see details in Sec. II C) is also included as a reference for comparison. As discussed in

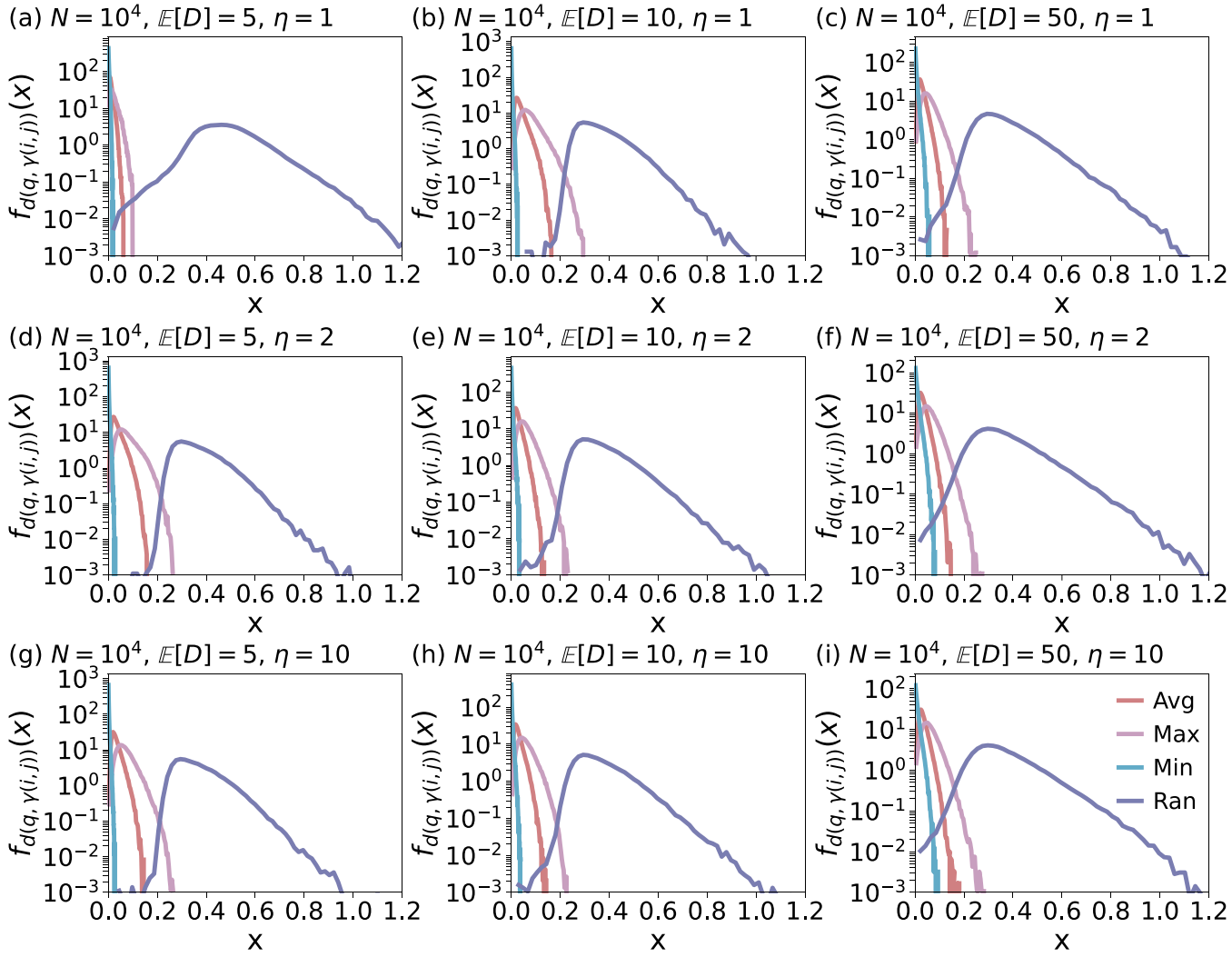


FIG. 11. The distribution of the average (avg), maximum (max), and minimum (min) distances between shortest path nodes and geodesics connecting shortest path end points in Waxman SRGGs with the connection probability function given by Eq. (4). For comparison, we also plot the distribution of distances to geodesic from randomly selected nodes (ran). The network size $N = 10^4$ and the parameter $\beta = 1$. For each expected degree $\mathbb{E}[D]$ and parameter η, we generated an SRGG and considered 10^6 random shortest paths.

Sec. II C, the RGG+net approach reconstructs a single deterministic graph for a given network size N, expected degree $\mathbb{E}[D]$, and node coordinates $\mathbf{x}_i$ and $\mathbf{y}_i$. Consequently, this method yields only a single operating point, rather than a curve obtained by varying a decision threshold or model parameter, and is therefore not included in the AUPR analysis in this Appendix.

Consistent with the precision results shown in Fig. 5, the AUPR of the distance-to-geodesic method for identifying shortest path nodes is significantly higher than expected by chance. Moreover, the performance of the distance-to-geodesic approach is less sensitive to noise compared to the SRGG reconstruction method.

APPENDIX F: APPROXIMATION OF THE STRETCH

In this Appendix, we examine the assumptions and validity of the mean-field-type approximation for the average shortest path stretch, in which the stretch is approximated as the product of the average edge length and the average hop count [see Eq. (10)]. Figure 15 compares the exact average stretch with this mean-field-type approximation as a function of $\langle D \rangle$, for three representative values of β [$\beta = 2.5$, $\beta = 64$, and $\beta = 128$, shown in panels (a)–(c), respectively]. The approximation shows good quantitative agreement in the large-β regime corresponding to the RGG limit, as evidenced by the close overlap of the two curves both in panels (b) and (c). For small values of β, such as $\beta = 2.5$, the agreement becomes less accurate, although the approximation still captures the qualitative behavior, particularly in the subcritical regime (i.e., for $\langle D \rangle$ below the percolation threshold). Perhaps more importantly, in each panel the approximation reproduces the overall trend of the exact stretch, including the characteristic nonmonotonic dependence on $\langle D \rangle$. This indicates that the dominant mechanisms governing the stretch are already

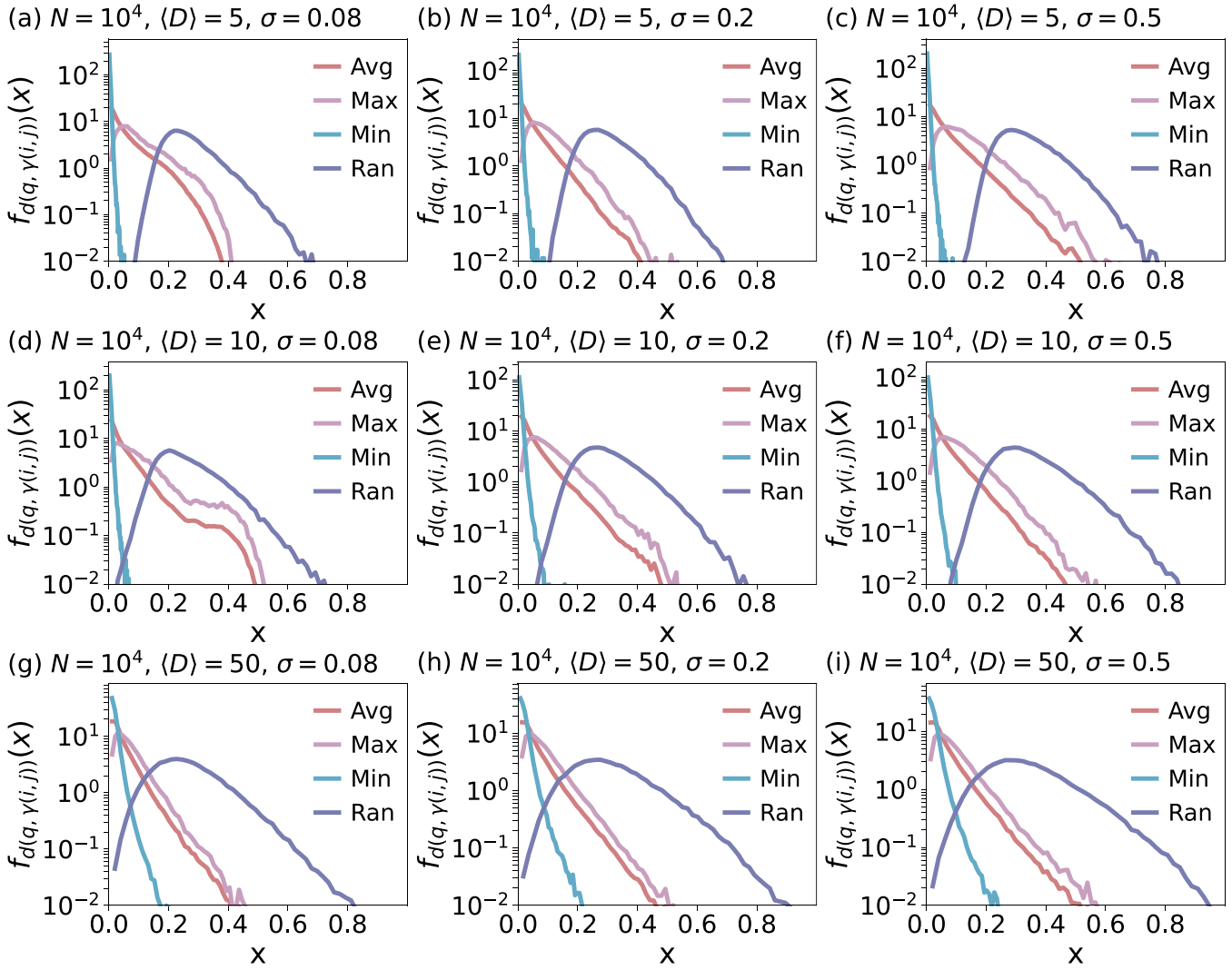


FIG. 12. The distribution of the average (avg), maximum (max), and minimum (min) distances between shortest path nodes and geodesics connecting shortest path end points in SRGGs where nodes are *not* uniformly scattered and the connection probability function is given by Eq. (5). The distribution of distances to geodesic from randomly selected nodes (ran) is also presented. The network size $N = 10^4$. For each average degree $\langle D \rangle$ and standard deviation σ of the Gaussian distribution used to sample node coordinates, we generated an SRGG and considered 10^6 random shortest paths.

captured within the mean-field framework, despite its simplifying assumptions.

The observed deviations between the exact and approximate results arise from several sources. First, quantities such as hop count and edge lengths along shortest paths are inherently correlated. Furthermore, shortest paths themselves emerge from a global optimization procedure, which systematically biases the selection toward atypically short or geometrically favorable edges. These effects introduce nontrivial statistical dependencies that render an exact analytical treatment of the stretch practically infeasible.

The mean-field approximation simplifies this picture by treating all node pairs as statistically equivalent and assuming that edge lengths along shortest paths are homogeneous and representative of the overall link-length distribution. Formally, this corresponds to replacing the conditional expectation $\langle d_l \,|\, l \rangle$ with the unconditional mean $\langle d_l \rangle$, and neglecting correlations between hop count and edge lengths. Under these assumptions, the average stretch factorizes as $\langle S \rangle \approx \langle d_l \rangle \langle h \rangle$, as given in Eq. (10). Although highly simplified, this approach still enables an analytical treatment of the stretch, which would otherwise be intractable. The accuracy of the approximation, however, depends strongly on β. In the large-β limit, edge lengths are narrowly distributed and variability along shortest paths is limited, reducing correlations and improving the validity of the mean-field assumption. In contrast, for small β, the broader distribution of edge lengths enhances correlations, leading to larger deviations. This behavior is consistent with the trends observed in the panels of Fig. 15.

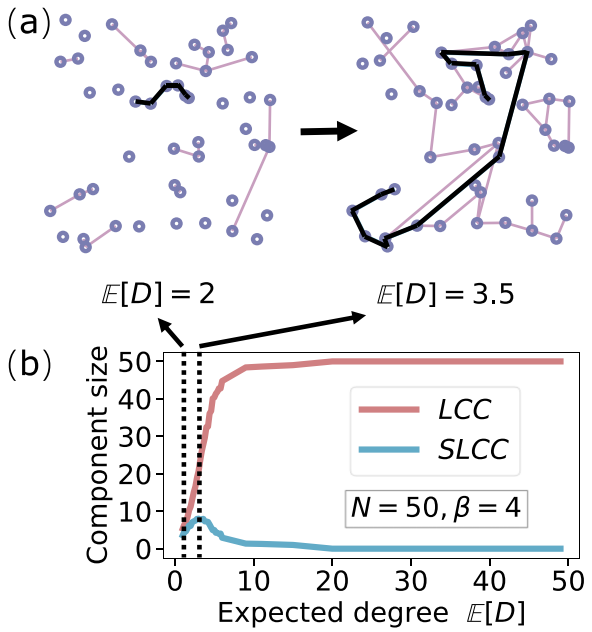


FIG. 13. (a) Visualizations of toy SRGGs with $N = 50$ nodes and inverse temperature $\beta = 4$ for increasing expected degree, from left to right: $\mathbb{E}[D] = 2$, $\mathbb{E}[D] = 3.5$. In each graph, a representative shortest path is highlighted in black. (b) Average sizes of the largest connected component (LCC) and the second largest connected component (SLCC) for SRGGs with $N = 50$ and $\beta = 4$ as a function of the expected degree $\mathbb{E}[D]$. The emergence of a giant connected component occurs at the critical expected degree $D_c = 3.5$, identified by the maximum of the SLCC. All LCC and SLCC values are averaged over 100 independent realizations. Accordingly, the graph visualizations in panel (a) correspond to the disconnected and critically connected regimes.

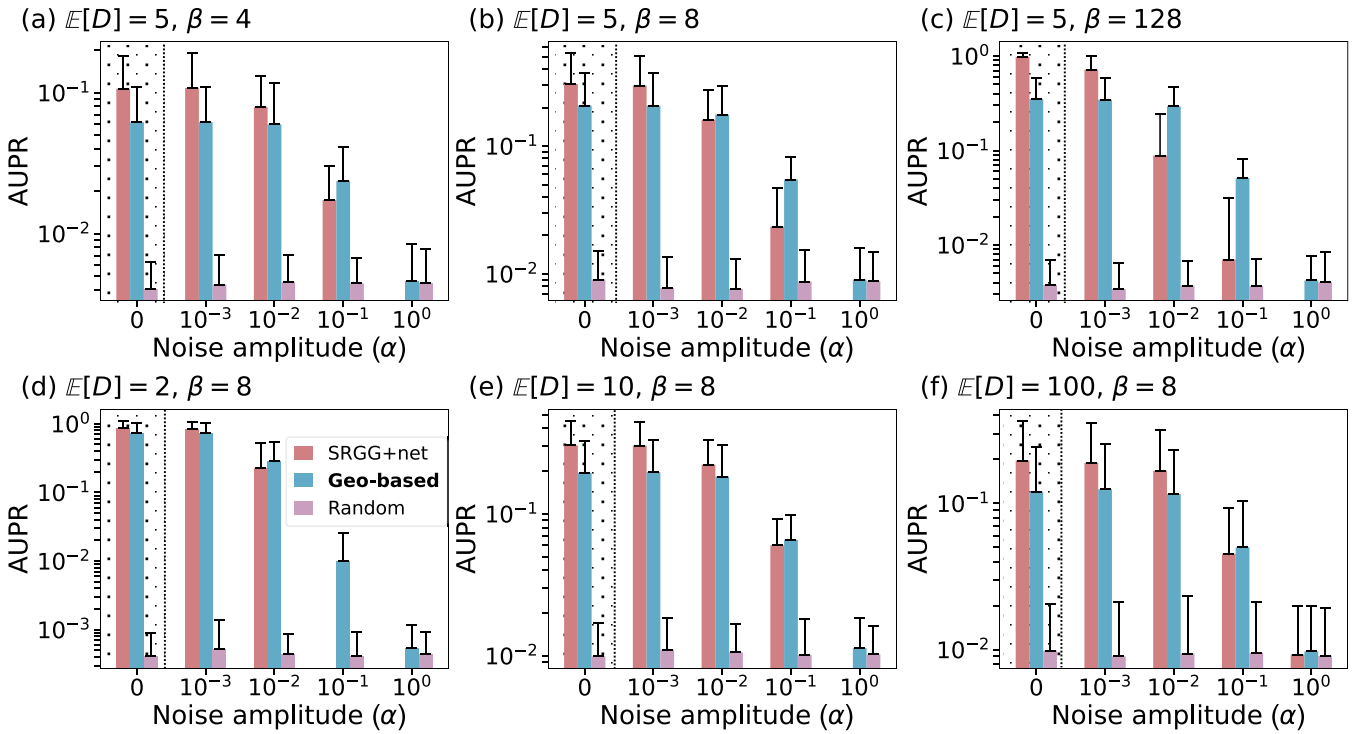


FIG. 14. AUPR of shortest path identification using the distance-to-geodesic (geometry-based) method. Panels show the average AUPR as a function of the coordinate noise amplitude α, for different combinations of $\mathbb{E}[D]$ and β. All experiments are performed on SRGG networks with $N = 10^4$ nodes. Each bar represents the average AUPR over 10^2 randomly selected node pairs and the error bars indicate standard deviations.

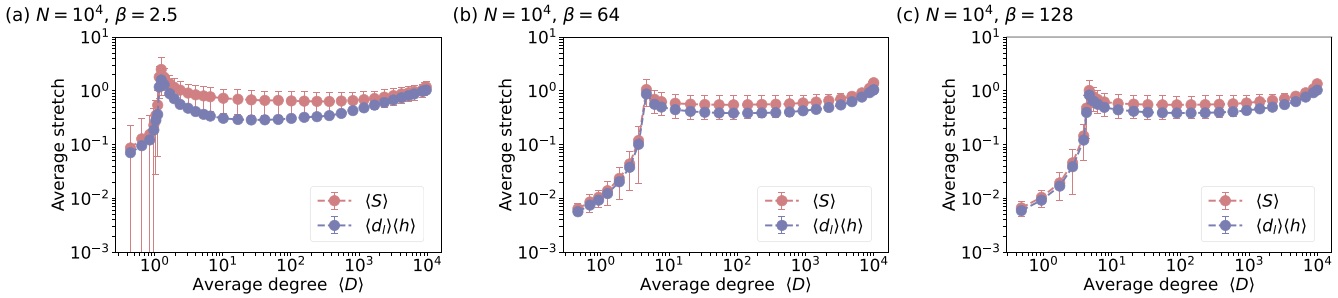


FIG. 15. Comparison of the average stretch and the product of the average link length and average hop count as functions of the average degree for SRGG networks with $N = 10^4$ nodes, for different values of $\boldsymbol{\beta}$. The average stretch is computed over 10^6 randomly selected node pairs, and the error bars indicate standard deviations.

[1] Z. Wang and J. Crowcroft, Analysis of shortest-path routing algorithms in a dynamic network environment, Comput. Commun. Rev. **22**, 63 (1992).

[2] C. W. Ahn and R. S. Ramakrishna, A genetic algorithm for shortest path routing problem and the sizing of populations, IEEE Trans. Evol. Comput. **6**, 566 (2002).

[3] D. Bertsekas, Dynamic behavior of shortest path routing algorithms for communication networks, IEEE Trans. Autom. Control **27**, 60 (1982).

[4] D. B. Johnson, Efficient algorithms for shortest paths in sparse networks, J. ACM **24**, 1 (1977).

[5] R. M. Gomathi and J. M. L. Manickam, Energy efficient shortest path routing protocol for underwater acoustic wireless sensor network, Wireless Pers. Commun. **98**, 843 (2018).

[6] L. Fu, D. Sun, and L. R. Rilett, Heuristic shortest path algorithms for transportation applications: State of the art, Comput. Oper. Res. **33**, 3324 (2006).

[7] S. Pallottino and M. G. Scutella, Shortest path algorithms in transportation models: Classical and innovative aspects, in *Equilibrium and Advanced Transportation Modelling*, edited by M. Patrice and N. Sang (Springer, Boston, MA, 1998), pp. 245–281.

[8] D. Van Vliet, Improved shortest path algorithms for transport networks, Transp. Res. **12**, 7 (1978).

[9] D. Zhang, Y. Shou, and J. Xu, A MapReduce-based approach for shortest path problem in road networks, J. Ambient Intell. Humaniz. Comput. **15**, 1251 (2024).

[10] W.-B. Du, Z.-X. Wu, and K.-Q. Cai, Effective usage of shortest paths promotes transportation efficiency on scale-free networks, Physica A **392**, 3505 (2013).

[11] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. **6**, 888 (2010).

[12] N. E. Friedkin, Theoretical foundations for centrality measures, Am. J. Sociol. **96**, 1478 (1991).

[13] C. F. Moukarzel, Spreading and shortest paths in systems with sparse long-range connections, Phys. Rev. E **60**, R6263(R) (1999).

[14] S. Pei and H. A. Makse, Spreading dynamics in complex networks, J. Stat. Mech. (2013) P12002.

[15] R. K. Karunakaran, S. Manuel, and E. N. Satheesh, Spreading information in complex networks: An overview and some modified methods, in *Graph Theory—Advanced Algorithms and Applications*, edited by B. Sirmacek (IntechOpen, London, 2017), Chap. 1.

[16] M. Kitsak, A. Ganin, A. Elmokashfi, H. Cui, D. A. Eisenberg, D. L. Alderson, D. Korkin, and I. Linkov, Finding shortest and nearly shortest path nodes in large substantially incomplete networks by hyperbolic mapping, Nat. Commun. **14**, 186 (2023).

[17] B. Y. Chen, W. H. K. Lam, A. Sumalee, and Z.-l. Li, Reliable shortest path finding in stochastic networks with spatial correlated link travel times, Int. J. Geogr. Inf. Sci. **26**, 365 (2012).

[18] E. W. Dijkstra, A note on two problems in connexion with graphs, Numer. Math. **1**, 269 (1959).

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 4th ed. (MIT Press, Cambridge, MA, 2022).

[20] R. Bellman, On a routing problem, Q. Appl. Math. **16**, 87 (1958).

[21] A. Schrijver, On the history of the shortest path problem, Doc. Math. **17**, 155 (2012).

[22] L. R. Ford, *Network Flow Theory* (Rand Corporation, Santa Monica, CA, 1956).

[23] K. Luck *et al.*, A reference map of the human binary protein interactome, Nature (London) **580**, 402 (2020).

[24] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha, A survey of COVID-19 contact tracing apps, IEEE Access **8**, 134577 (2020).

[25] K. C. Claffy and D. Clark, Challenges in measuring the Internet for the public interest, J. Inf. Policy **12**, 195 (2022).

[26] L. Lü and T. Zhou, Link prediction in complex networks: A survey, Physica A **390**, 1150 (2011).

[27] R. Guimerà and M. Sales-Pardo, Missing and spurious interactions and the reconstruction of complex networks, Proc. Natl. Acad. Sci. USA **106**, 22073 (2009).

[28] R. R. Nadakuditi and M. E. J. Newman, Graph spectra and the detectability of community structure in networks, Phys. Rev. Lett. **108**, 188701 (2012).

[29] C. Bordenave, M. Lelarge, and L. Massoulié, Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs, Ann. Probab. **46**, 1 (2018).

[30] E. Abbe, Community detection and stochastic block models, Found. Trends Commun. Inf. Theory **14**, 1 (2018).

[31] M. MacMahon and D. Garlaschelli, Community detection for correlation matrices, Phys. Rev. X **5**, 021006 (2015).

[32] J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-aron, J. Sijbers, and E. Fieremans, Denoising of diffusion MRI using random matrix theory, NeuroImage **142**, 394 (2016).

[33] S. Bartolucci, F. Caccioli, F. Caravelli, and P. Vivo, Distribution of centrality measures on undirected random networks via the cavity method, Proc. Natl. Acad. Sci. USA **121**, e2403682121 (2024).

[34] M. Boguñá, I. Bonamassa, M. De Domenico, S. Havlin, D. Krioukov, and M. Á. Serrano, Network geometry, Nat. Rev. Phys. **3**, 114 (2021).

[35] A. P. Kartun-Giles, M. Barthelemy, and C. P. Dettmann, Shape of shortest paths in random spatial networks, Phys. Rev. E **100**, 032315 (2019).

[36] J. Díaz, D. Mitsche, G. Perarnau, and X. Pérez-Giménez, On the relation between graph distance and Euclidean distance in random geometric graphs, Adv. Appl. Probab. **48**, 848 (2016).

[37] R. B. Ellis, J. L. Martin, and C. Yan, Random geometric graph diameter in the unit ball, Algorithmica **47**, 421 (2007).

[38] M. Bradonjić, R. Elsässer, T. Friedrich, T. Sauerwald, and A. Stauffer, Efficient broadcast on random geometric graphs, in *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms* (SIAM, Austin, Texas, 2010), pp. 1412–1421.

[39] T. Friedrich, T. Sauerwald, and A. Stauffer, Diameter and broadcast time of random geometric graphs in arbitrary dimensions, Algorithmica **67**, 65 (2013).

[40] F. Voss, C. Gloaguen, F. Fleischer, and V. Schmidt, Distributional properties of Euclidean distances in wireless networks involving road systems, IEEE J. Sel. Areas Commun. **27**, 1047 (2009).

[41] M. Raftopoulou, R. Litjens, and P. Van Mieghem, Fréchet distribution in geometric graphs for drone networks, Phys. Rev. E **106**, 024301 (2022).

[42] R. Hekmat and P. Van Mieghem, Connectivity in wireless ad-hoc networks with a log-normal radio model, Mobile Networks Appl. **11**, 351 (2006).

[43] Y. Zhang, H. Zhang, W. Sun, and C. Pan, Connectivity analysis for vehicular ad hoc network based on the exponential random geometric graphs, in *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings* (IEEE, Dearborn, MI, 2014), pp. 993–998.

[44] B. Boyacı, T. H. Dang, and A. N. Letchford, Vehicle routing on road networks: How good is Euclidean approximation? Comput. Oper. Res. **129**, 105197 (2021).

[45] S. Buczkowska, N. Coulombel, and M. de Lapparent, A comparison of Euclidean distance, travel times, and network distances in location choice mixture models, Networks Spat. Econ. **19**, 1215 (2019).

[46] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, Covert cyber assault detection in smart grid networks utilizing feature selection and Euclidean distance-based machine learning, Appl. Sci. **8**, 772 (2018).

[47] U. Islambekov, A. K. Dey, Y. R. Gel, and H. V. Poor, Role of local geometry in robustness of power grid networks, in *Proceedings of the 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (IEEE, Anaheim, CA, 2018), pp. 885–889.

[48] C. Sommer, Shortest-path queries in static networks, ACM Comput. Surv. **46**, 1 (2014).

[49] R. W. van der Hofstad, *Random Graphs and Complex Networks: Volume 1*, Cambridge Series in Statistical and Probabilistic Mathematics Vol. 43 (Cambridge University Press, Cambridge, 2017).

[50] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, 2003), Vol. 5.

[51] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, Phys. Rev. E **82**, 036106 (2010).

[52] G. García-Pérez, M. Boguñá, and M. Á. Serrano, Multiscale unfolding of real networks by geometric renormalization, Nat. Phys. **14**, 583 (2018).

[53] J. Chellig, N. Fountoulakis, and F. Skerman, The modularity of random graphs on the hyperbolic plane, J. Complex Networks **10**, cnab051 (2021).

[54] S. G. Balogh, B. Kovács, and G. Palla, Maximally modular structure of growing hyperbolic networks, Commun. Phys. **6**, 76 (2023).

[55] S. G. Balogh and G. Palla, Intra-community link formation and modularity in ultracold growing hyperbolic networks, Physica A **642**, 129784 (2024).

[56] S. G. Balogh, B. Sulyok, T. Vicsek, and G. Palla, CLOVE, a travelling salesman's approach to hyperbolic embeddings of complex networks with communities, Commun. Phys. **8**, 397 (2025).

[57] C. V. Cannistraci and A. Muscoloni, Geometrical congruence, greedy navigability and myopic transfer in complex networks and brain connectomes, Nat. Commun. **13**, 7308 (2022).

[58] C. D. Howard and C. M. Newman, Geodesics and spanning trees for Euclidean first-passage percolation, Ann. Probab. **29**, 577 (2001).

[59] A. Chassein, T. Dokka, and M. Goerigk, Algorithms and uncertainty sets for data-driven robust shortest path problems, Eur. J. Oper. Res. **274**, 671 (2019).

[60] M. D. Penrose, Connectivity of soft random geometric graphs, Ann. Appl. Probab. **26**, 986 (2016).

[61] G. Mao and B. D. O. Anderson, Connectivity of large wireless networks under a general connection model, IEEE Trans. Inf. Theory **59**, 1761 (2013).

[62] K. Bringmann, R. Keusch, and J. Lengler, Geometric inhomogeneous random graphs, Theor. Comput. Sci. **760**, 35 (2019).

[63] T. Bläsius, T. Friedrich, M. Katzmann, U. Meyer, M. Penschuck, and C. Weyand, Efficiently generating geometric inhomogeneous and hyperbolic random graphs, Network Sci. **10**, 361 (2022).

[64] J. Dall and M. Christensen, Random geometric graphs, Phys. Rev. E **66**, 016121 (2002).

[65] P. Van Mieghem, Paths in the simple random graph and the Waxman graph, Probab. Eng. Inf. Sci. **15**, 535 (2001).

[66] B. M. Waxman, Routing of multipoint connections, IEEE J. Sel. Areas Commun. **6**, 1617 (1988).

[67] V. Erba, S. Ariosto, M. Gherardi, and P. Rotondo, Random geometric graphs in high dimension, Phys. Rev. E **102**, 012306 (2020).

[68] M. Boguñá, D. Krioukov, P. Almagro, and M. Á. Serrano, Small worlds and clustering in spatial networks, Phys. Rev. Res. **2**, 023040 (2020).

[69] D. M. W. Powers, Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation, arXiv:2010.16061.

[70] Ž. Vujović, Classification model evaluation metrics, Int. J. Adv. Comput. Sci. Appl. **12**, 599 (2021).

[71] J. Davis and M. Goadrich, The relationship between precision-recall and ROC curves, in *Proceedings of the 23rd International Conference on Machine Learning*, *ICML'06* (ACM Press, New York, USA, 2006), pp. 233–240.

[72] P. Van Mieghem, *Performance Analysis of Complex Networks and Systems* (Cambridge University Press, Cambridge, UK, 2014).

[73] G. Hooghiemstra and P. Van Mieghem, On the mean distance in scale free graphs, Methodol. Comput. Appl. Probab. **7**, 285 (2005).

[74] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, A universal model for mobility and migration patterns, Nature (London) **484**, 96 (2012).

[75] A. Allard, M. Á. Serrano, G. García-Pérez, and M. Boguñá, The geometric nature of weights in real complex networks, Nat. Commun. **8**, 14103 (2017).

[76] I. Voitalov, P. Van Der Hoorn, M. Kitsak, F. Papadopoulos, and D. Krioukov, Weighted hypersoft configuration model, Phys. Rev. Res. **2**, 043157 (2020).

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

[78] Y. Yang, T. Nishikawa, and A. E. Motter, Small vulnerable sets determine large network cascades in power grids, Science **358**, eaan3184 (2017).

[79] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, Vital nodes identification in complex networks, Phys. Rep. **650**, 1 (2016).

[80] Z. Qiu, T. Fan, M. Li, and L. Lü, Identifying vital nodes by Achlioptas process, New J. Phys. **23**, 033036 (2021).

[81] L. C. Freeman, A set of measures of centrality based on betweenness, Sociometry **40**, 35 (1977).

[82] M. Kitsak, S. Havlin, G. Paul, M. Riccaboni, F. Pammolli, and H. E. Stanley, Betweenness centrality of fractal and non-fractal scale-free model networks and tests on real networks, Phys. Rev. E **75**, 056115 (2007).