Betweenness centrality in dense spatial networks

Vincent Verbavatz [*]

Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191, Gif-sur-Yvette, France
and École des Ponts ParisTech, F-77420 Champs-sur-Marne, France

Marc Barthelemy [†]

Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191, Gif-sur-Yvette, France
and CAMS (CNRS/EHESS) 54 Boulevard Raspail, 75006 Paris, France

The betweenness centrality (BC) is an important quantity for understanding the structure of complex large networks. However, its calculation is in general difficult and known in simple cases only. In particular, the BC has been exactly computed for graphs constructed over a set of N points in the infinite density limit, displaying a universal behavior. We reconsider this calculation and propose an expansion for large and finite densities. We compute the lowest nontrivial order and show that it encodes how straight are shortest paths and is therefore nonuniversal and depends on the graph considered. We compare our analytical result to numerical simulations obtained for various graphs such as the minimum spanning tree, the nearest neighbor graph, the relative neighborhood graph, the random geometric graph, the Gabriel graph, or the Delaunay triangulation. We show that in most cases the agreement with our analytical result is excellent even for densities of points that are relatively low. This method and our results provide a framework for understanding and computing this important quantity in large spatial networks.

I. INTRODUCTION

There are many centralities for characterizing the importance of a node in a network [1]. Among those, local quantities fail to give interesting information about global structures while path-related measures are more relevant to describe the large-scale organization of networks. In particular, the betweenness centrality (BC), introduced in Ref. [2] is a good probe of the structure of a network. Also, if one assumes that (i) individuals or goods travel on shortest paths in the network, and (ii) the demand is uniform (each pair of nodes constitutes an origin-destination couple), then the BC of a node (or an edge) corresponds to the local traffic that can be found at this node (or edge). In reality, the two assumptions are not always satisfied and how much of the real traffic the BC can explain is a debated question [3–5]. In general, highly congested points are signaled by very large values of the BC and this is relevant not only for transportation networks but also for communication networks such as the Internet where information packets can experience congestion problems at routers. In a router-based communication network, all nodes are connected to it directly and the BC is irrelevant in this case. A new direction for modern design of physical layer networks is to construct "wireless ad hoc networks," where routers are absent and packets of information are routed in a multihop fashion between any two nodes [6–8]. This design allows for

much larger and flexible networks and are nowadays realized under Wi-Fi direct standards. For these decentralized systems, the BC is a very relevant quantity and can be used as a criteria for identifying cluster nodes [9] or to identify the vulnerability backbone of the network [10]. Still in communication networks, it is intuitive to think that the traffic between nodes tends to go through a small core of nodes. In this case, the shortest paths are somehow curved inwards and it has been suggested that this is related to the global curvature of the network [11,12]. A natural way to measure the impact of the structure on the load in the network is then to understand how the maximum traffic—approximated by the maximum BC—depends on various graph properties and scales with the system size measured by the number of nodes [11]. Narayam and Saniee [11] studied empirically various networks and found essentially two families characterized by different values of the exponent that governs this scaling. These authors proposed the idea that this behavior is controlled by the curvature of the network and this was justified mathematically by Jonckheere *et al.* [12].

The BC is of interest for spatial networks (planar or not) and real-world applications such as transport networks [13–16]. For street networks in cities, studies unveiled the presence of crucial nodes with very large BC [13,16,17] and the localization of these congested points can reveal interesting features about the organization of the network [16,17] and its large-scale organization [14,16,18]. More recently, an empirical study on almost 100 cities worldwide demonstrated that the empirical BC distribution seems to be an invariant in world cities [19]: its structure results from the superimposition

———————
[*]vincent.verbavatz@ipht.fr
[†]marc.barthelemy@ipht.fr

of a backbone tree (corresponding to the minimum spanning tree) and redundant streets in agreement with the picture proposed in Ref. [20] for synthetic networks. The interest in the BC also lies in the fact that it is correlated with some economic features such as the density of retail stores [21–26] (we note that another analysis about Buenos Aires challenges this relation [27]).

From a more theoretical point of view, however, few general results are known about the BC [28–30]. Particular geometries such as branches and loops are well understood [31] while more involved geometries were studied only recently [32]. Yet, a major theoretical result concerns planar graphs constructed on random points embedded in a bounded set. In this generic case, when there is an infinite number of points in the domain (usually a square or a disk), the shortest paths on the graph will likely be straight lines and the BC of any point can be computed exactly as a function of its position [33]. We note here that how the shortest path deviate from the straight line is interesting itself [34] and more generally, the shape of shortest paths is an important problem [35] in relation with the first passage percolation problem, a well-known subject in statistical physics (see, for example, Ref. [36] and references therein). Here, we extend this infinite density calculation to the case of large but finite densities and the organization of this paper is as follows. We will first define the BC and recall some of its general properties and results (in particular for simple graphs). We will then present the perturbation expansion around the infinite density limit and test these results for various graphs constructed over a set of points in the plane.

II. THE BETWEENNESS CENTRALITY

A. Definition and generalities

The betweenness centrality for a node i in a graph G with N nodes is defined as [2]

$$g(i) = \frac{1}{\mathcal{N}} \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}, \tag{1}$$

where σ_{st} is the number of shortest paths from node s to node t and $\sigma_{st}(i)$ the number of these shortest paths that go through node i. The quantity $\mathcal{N}$ is a normalization that we choose here $\mathcal{N} = (N-1)(N-2)$ so that the BC is in [0,1]. We can define in a similar way the BC $g(e)$ for an edge e using the quantity $\sigma_{st}(e)$ which is the number of shortest path from node s to node t going through the link e.

It can be shown that the BC averaged over all nodes $\bar{g} = 1/N \sum_i g(i)$ is proportional to the average shortest path ℓ [28,29] (which is the shortest distance between two nodes in the graph, averaged over all pairs of nodes). This allows in particular to understand that adding a link to the graph will decreases the average BC. More precisely, it can be shown [29] that if we add to a graph of size N a link of shortest path length d, the average BC transforms according to

$$\bar{g} \to \bar{g} - \frac{2(d-1)}{N}. \tag{2}$$

We note here that if the BC decreases on average, it does not imply that the BC of all nodes decreases when adding new

links. Locally, we can observe an increase of the BC of some points.

B. One and two dimensional grids

For one-dimensional lattices with N nodes, it is easy to show that the BC of node i ($i \in [\![1, N]\!]$) is given by

$$g(i) = \frac{i}{N}\left(1 - \frac{i}{N}\right). \tag{3}$$

The barycenter of all nodes $i_b = N/2$ is then also the most central node. There are other results available in 1D and we mention here the example of the random geometric graph [37].

For a two-dimensional square grid, it is easy to express the BC of a node as a sum of combinatorial factors that count the number of paths. The number of paths between points (a, b) and (i, j) with $a < i$ and $b < j$ being $\binom{i+j-a-b}{i-a}$, the centrality of the node (i, j) on the grid $[\![-L, L]\!] \times [\![-L, L]\!]$ is

$$g(i, j) = \frac{1}{4L^2} \sum_{\sigma \in \{-1,1\}} \sum_{a=-L}^{i} \sum_{b=-L}^{j}$$
$$\times \sum_{c=i}^{L} \sum_{d=j}^{L} \frac{\binom{j-b+\sigma(i-a)}{\sigma(i-a)}\binom{d-j+\sigma(c-i)}{\sigma(c-i)}}{\binom{c+d-a-b}{\sigma(c-a)}}, \tag{4}$$

where $\sigma = \pm 1$ corresponds to nodes with $j < 0$ and $j > 0$, respectively. This expression is difficult to analyze, but we can resort to the simple approximation described in Fig. 1(a). We assume that the number of paths going through the node $(i, 0)$ is proportional to the product of areas described in gray in Fig. 1, normalized by the total number of paths (we have to multiply the result by a factor 2 by symmetry). We thus obtain

$$g(i, 0) \propto L(L - i) \times (i + L)L/L^4 = 1 - \left(\frac{i}{L}\right)^2. \tag{5}$$

We compare this approximation to the exact numerical result showing a very good agreement [Fig. 1(b)]. The discrepancy appears essentially for $i \approx L$ where the approximation predicts $g(L) = 0$ which is exact for $L \gg 1$. Using this same argument, we find that for any node located at (i, j), the BC reads

$$g(i, j) \simeq (1 - i^2)(1 - j^2). \tag{6}$$

We compare this expression to the exact result and show in Fig. 1(c) the relative error. Here also the most important errors appear at the boundary of the square but, despite its simplicity, the approximation is very good in the bulk of the square.

C. Loops, branches, and more complex graphs

More complex graphs have also been studied from the perspective of the BC. In particular, in Ref. [31] a toy model made of a star network with N_b branches of size n and links of weight 1, superimposed to a loop at distance ℓ from the center and with links of weight w was considered. The BC at the center and on the loop were computed and it can be shown that the loop can be more central than the center if $w < w_c$ where the threshold scales as $w_c \sim n/N_b$. This sheds some light on empirical results about road networks where ring roads can be more central than the spatial center of the system.
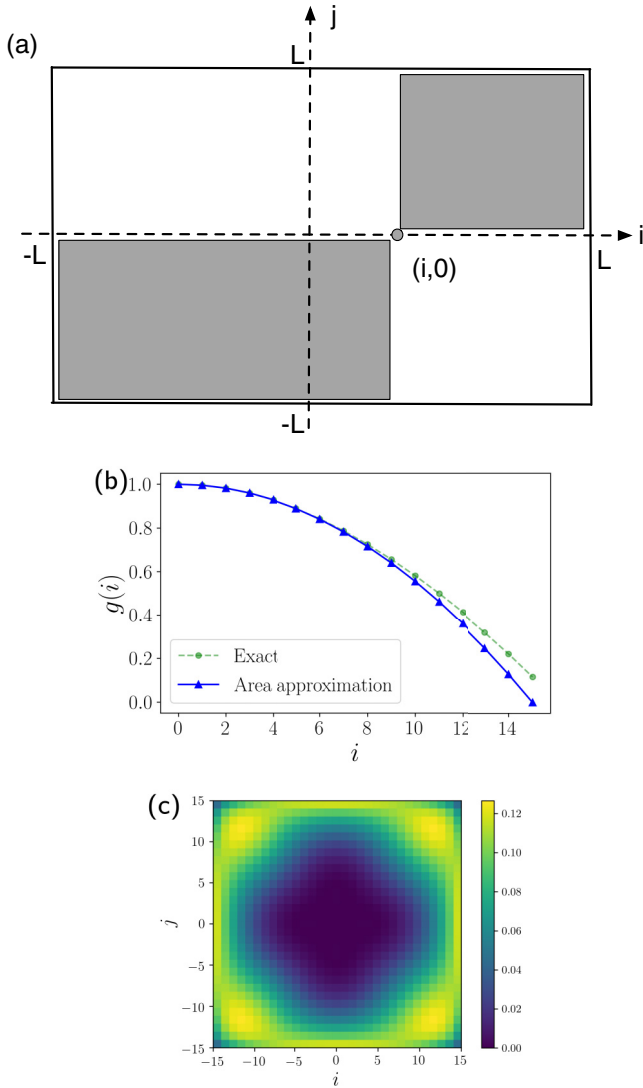
FIG. 1. (a) Approximation for computing the BC for the 2D grid: the BC at node $(i, 0)$ is proportional to the product of shaded areas (up to a factor two accounting for the white squares). (b) Comparison of the approximation of Eq. (5) to the analytical formula of Eq. (4) computed numerically for $L = 15$. (c) Relative error of the approximation of Eq. (6) on the square grid.

Also recently, a more complex structure was considered in Ref. [32]. In this study, the authors introduce a family of planar graphs composed by a square grid connected to an arboreal periphery. The BC was then computed at the center of the grid and some other important points such as the ones connecting the square grid to the peripheral trees.

In general, for complex networks that are not planar, the BC is increasing with the degree of the node as a power law $g(k) \sim k^\eta$ with an exponent η that is in general less than 2. The main reason for this bound is that the number of possible paths through a node of degree k is $k(k-1)$ which scales as k^2 for large k. This is then what we would obtain for the BC if all neighbors lead to regions with roughly the same number of nodes. If it is not the case, then some neighbors will be more

important than others and then not all the $k(k-1)$ paths are important and therefore $\eta < 2$.

D. Distribution of the BC in planar graphs

The distribution $P(g)$ of the BC was discussed recently in Ref. [19] where it was shown on almost 100 different road networks for worldwide cities that it is invariant. This invariance is a consequence of a bimodal regime where the high BC nodes belong to the underlying tree structure of the graph, and the low BC nodes to loops that provide alternate paths. More precisely, if we rescale the centrality by the number N of nodes $\tilde{g} = g/N$, then we obtain the invariant distribution as

$$P(\tilde{g}) \sim \frac{e^{-\tilde{g}/\gamma}}{\tilde{g}^\alpha}, \qquad (7)$$

where the exponent $\alpha \approx 1$ can be explained with a simple tree model [19] and where γ depends on the specific graph. This invariance in particular suggests that the interesting information about the BC lies not in its statistical properties but rather in its spatial distribution, and where the high BC nodes are located [16,19] which depends in general on details of the structure of the graph.

III. BETWEENNESS CENTRALITY IN DENSE AND QUASIDENSE GRAPHS

A. Graphs constructed over a set of points

We will consider here different graphs constructed over a set of points in a bounded domain. We consider a Poisson process where N points are distributed randomly in a plane domain $\mathcal{D}$ of area V (which will be a disk or a square). The density of nodes is denoted by $\rho = N/V$. There are multiple ways to connect these points to each other and we will consider here various graphs.

First, we will consider graphs that are constructed by connecting a node to its k-nearest neighbors (k-NN with $k = 7$), the random geometric graph (RGG) that connects points closer than a threshold distance d (we choose $d = 2/\sqrt{\rho}$, which is the typical distance for which the graph is connected), the minimal spanning tree (MST) that connects all the vertices together, without any cycles and with the minimum possible total edge weight, the Delaunay triangulation (DT) that gives a triangulation such that no point is inside the circumcircle of any triangle of the triangulation, the Gabriel graph (GG) which is the subgraph of DT where any two distinct points P and Q are adjacent precisely when the closed disk having PQ as a diameter contains no other points, and finally the relative neighborhood graph (RNG) which connects two points P and Q by an edge whenever there does not exist a third point R that is closer to both P and Q than they are to each other. These graphs represent many important cases and are widely studied in mathematics (see, for example, Refs. [34,38,39]), in statistical physics (see Ref. [28] and references therein), or for wireless network applications [40]. Understanding the BC for these cases thus represents an important step toward a general theory of the BC in spatial networks.

B. Perturbation around the infinite density

We compute the BC of nodes in $\mathcal{D}$ in the quasidense limit $(1 \ll \rho < \infty)$. We aim at finding an expression for the BC that depends on the absolute position of points in the plane only and not on the specific graph. In Ref. [33], it was shown that in the dense limit $(\rho = \infty)$ on a disk, the BC of a node depends on its distance to the center only, whatever the specific graph. This approximation relies on the fact that shortest paths in this limit are essentially straight lines, which explains the universality of the result. However, for finite densities, the shortest paths display significant transversal deviations and we expect nonuniversal corrections.

We present here a perturbation expansion at the lowest nontrivial order of this previous result when the density is finite. We denote by (i, j, κ) random nodes (among N nodes of a graph G) inside a disk domain $\mathcal{D}$ of area V. The quantity $SP(i, j) = \{x_{ij_1}, ..., x_{ij_m}, ..., x_{ij_n}\}$ denotes the shortest path between points i and j (that we assume to be unique, which for spatial networks is expected—a degeneracy would imply exactly the same distance between two nodes which is very unlikely, in contrast with the topological distance which is an integer that counts the number of jumps).

We define the indicator function

$$\sigma_{ij}(\kappa) = \mathbf{1}_{\kappa \in SP(i,j)} = \sum_m \mathbf{1}_{\kappa = x_{ij_m}}. \tag{8}$$

This indicator function $\sigma_{ij}(\kappa)$ is equal to unity if κ is in $SP(i, j)$ and zero otherwise. The betweenness centrality for the node κ is then

$$g(\kappa) = \frac{1}{2} \sum_i \sum_j \frac{\sigma_{ij}(\kappa)}{\sigma_{ij}}, \tag{9}$$

where i and j are nodes of the graph. For large ρ, we use a continuous approximation and write

$$\sigma_{ij}(\kappa) = \int_{t_i}^{t_j} dt \ \delta[\kappa - x(t)], \tag{10}$$

where the shortest path $\{x(t) \in SP(i, j)\}$ is parametrized by $t \in [t_1, t_2]$ where t_1 and t_2 correspond to the endpoints (δ is the Dirac δ function). The BC for κ is then

$$g(\kappa) = \frac{1}{2V^2} \int_{\mathcal{D}} d\mathbf{r}_i \int_{\mathcal{D}} d\mathbf{r}_j \ \sigma_{ij}(\kappa). \tag{11}$$

In the continuous limit [33] $\rho \to \infty$, the shortest paths are straight segments and the indicator reads as $\sigma_{ij} = \delta[x \cos(\phi) + y \sin(\phi) - p]$, where $\kappa = (x, y)$ and where the segment (i, j) is parametrized by p and ϕ (for this type of parametrization, see, for example, Ref. [41]). This means that κ is in $SP(i, j)$ if and only if κ is on the line between i and j. In particular, it is easy to check that $\int_{\mathcal{D}} d\kappa \ \sigma_{ij}(\kappa) = |t_2 - t_1|$ as expected.

In the quasidense limit $(1 \ll \rho < \infty)$, we define the average betweenness centrality for κ as the expectation the BC for κ,

$$\overline{g(\kappa)} = \mathbb{E}_G[g(\kappa)], \tag{12}$$

where $\mathbb{E}_G$ denotes the average over all the graphs (for a given connection rule) constructed over an ensemble of points
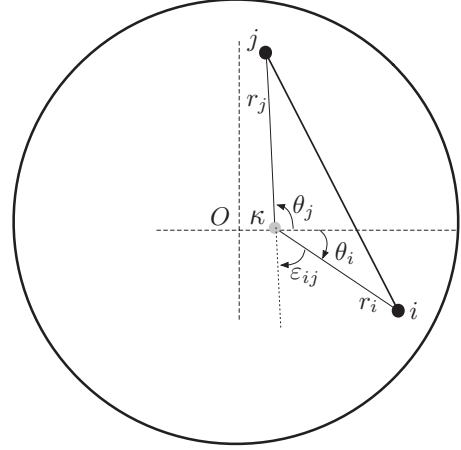

FIG. 2. Sketch of the system considered and notations. The origin of the polar system is κ and the nodes i and j have the coordinates (r_i, θ_i) and (r_j, θ_j) in this system. The deviation of the two segments $(i \kappa)$ and (κ, j) from the straight line is characterized by the angle $\varepsilon_{ij} = \theta_i - \theta_j + \pi$.

realizations. We then obtain

$$\overline{g(\kappa)} = \frac{1}{2V^2} \int_{\mathcal{D}} d\mathbf{r}_i \int_{\mathcal{D}} d\mathbf{r}_j \ \mathbb{E}_G[\sigma_{ij}(\kappa)]. \tag{13}$$

The quantity $\sigma_{ij}(\kappa)$ is an indicator function and its average is then a probability

$$\mathbb{E}_G(\sigma_{ij}(\kappa)) = \text{Prob}[\kappa \in SP(i, j)] \tag{14}$$

which we will denote by $\chi_{ij}(\kappa) = \text{Prob}[\kappa \in SP(i, j)]$. In the dense limit, shortest paths are straight lines and we have

$$\chi_{ij}(\kappa) = \delta[x \cos(\phi) + y \sin(\phi) - p]. \tag{15}$$

When the density is finite, the shortest paths deviate from the straight line and we define the angular deviation ϵ_{ij} from the straight line (i, j) in the frame of origin κ: $\epsilon_{ij} = \theta_i - \theta_j + \pi$ (see Fig. 2). Due to the statistical isotropy of the problem, it is enough to consider the node at distance κ from the center and at polar angle $\theta = 0$ and here and in the following we will work with the polar coordinate centered on this point.

We now express the probability $\chi_{ij}(\kappa)$ that κ is in $SP(i, j)$ for a given value of ϵ_{ij}, and the average BC can then formally be rewritten as the 5D integral

$$\overline{g(\kappa)} = \frac{1}{2V^2} \int_{\mathcal{D}^2} d\mathbf{r}_i d\mathbf{r}_j \int_0^\infty d\epsilon_{ij} \delta(\theta_i - \theta_j + \pi - \epsilon_{ij}) \chi_{ij}(\kappa|\epsilon_{ij}), \tag{16}$$

where $\chi_{ij}(\kappa|\epsilon_{ij})$ is the probability that $\{\kappa \in SP(i, j)\}$ conditioned by ϵ_{ij}. The δ function $\delta(\theta_i - \theta_j + \pi - \epsilon_{ij})$ ensures the definition of the angle ϵ_{ij}.

In the infinite density limit, we know from Ref. [33] that this conditional probability is given by

$$\chi_{ij}(\kappa|\epsilon_{ij}, \rho = \infty) \propto \left(\frac{1}{r_i} + \frac{1}{r_j} \right) \delta(\epsilon_{ij}), \tag{17}$$

and motivated by this case we assume the following generalization

$$\chi_{ij}(\kappa|\epsilon_{ij}) = \left(\frac{1}{r_i} + \frac{1}{r_j} \right) \chi(\epsilon_{ij}), \tag{18}$$

where $\chi(\epsilon)$ is an unknown function. Denoting ϵ_{ij} by ϵ, we assume that $\chi(\epsilon)$ is independent of (i, j, κ). This is a strong assumption that we empirically show to be correct for the k-NN, RGG, DT, and GG graphs [as we will further discuss below, this approximation is incorrect in the two cases of the MST and the RNG graphs and suggests that for these graphs the assumption that $\chi(\epsilon)$ is independent of (i, j, κ) is not correct]. Indeed, we show empirically that for these graphs, the function

$$\chi(\epsilon) = \frac{\chi_{ij}(\kappa|\epsilon)}{\left(\frac{1}{r_i} + \frac{1}{r_j}\right)} \tag{19}$$

can be fitted by a decreasing exponential function of ϵ with parameter ϵ_0,

$$\chi(\epsilon) = \epsilon_0(\rho)e^{-\epsilon/\epsilon_0(\rho)}, \tag{20}$$

where $\epsilon_0(\rho)$ is assumed to be a power law (see Fig. 3 and for the values of the fitting parameters for the different graphs, see Table I) of the form

$$\epsilon_0(\rho) = A\rho^{-\beta}. \tag{21}$$

Using the form of Eq. (18), we obtain

$$\overline{g(\kappa)} = \frac{1}{2V^2} \int_0^\infty d\epsilon\, \chi(\epsilon) \int_{\mathcal{D}} d\mathbf{r}_i \int_{\mathcal{D}} d\mathbf{r}_j$$

$$\times \left(\frac{1}{r_i} + \frac{1}{r_j}\right)\delta(\theta_i - \theta_j + \pi - \epsilon). \tag{22}$$

In the dense limit, we have $\chi(\epsilon) = \delta(\epsilon)$ and we recover the known result of Ref. [33]. To go beyond this infinite density result, we expand this function $\chi(\epsilon)$ around 0 to the second
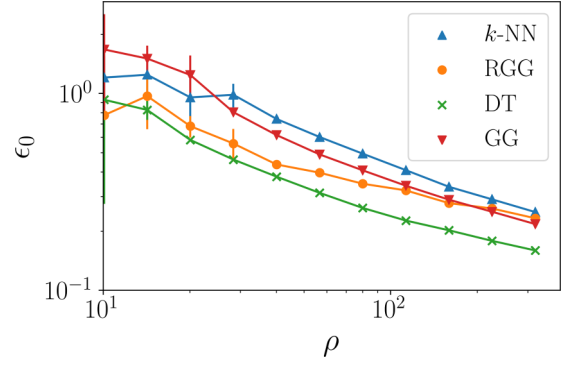

FIG. 3. Writing $\chi(\epsilon) = \epsilon_0(\rho)e^{-\epsilon/\epsilon_0(\rho)}$, we show that ϵ_0 is a smooth decreasing function of the density ρ, validating the shape of $\chi(\epsilon)$ for k-NN, RGG, DT, and GG graphs (the vertical error bars correspond to the dispersion). The graph suggests a power law relation of the form $\epsilon_0(\rho) \simeq A\rho^{-\beta}$ with $\beta \simeq 0.5 \pm 0.1$. We note that the exponent is not the exact same for all graphs, nor is the prefactor, thus making some types of graphs converging faster toward the dense regime than others.

order in ϵ_o

$$\chi(\epsilon) = \epsilon_0(\rho)e^{-\epsilon/\epsilon_0(\rho)}$$

$$\simeq \left[\delta(\epsilon) - \epsilon_0(\rho)\delta'(\epsilon) + \epsilon_0^2(\rho)\delta''(\epsilon)\right]. \tag{23}$$

Here, we used the distributional derivative of the Dirac δ function, which is defined so that for any compactly supported smooth test function ϕ, we have

$$\int dx\phi(x)\delta'(x) = -\int dx\phi'(x)\delta(x). \tag{24}$$

Inserting the expansion of Eq. (23) into the expression Eq. (22), we get

$$\overline{g(\kappa)} = \frac{1}{2V^2} \int_0^\infty d\epsilon\, \left[\delta(\epsilon) - \epsilon_0(\rho)\delta'(\epsilon) + \epsilon_0^2(\rho)\delta''(\epsilon)\right] \int_{\mathcal{D}} d\mathbf{r}_i \int_{\mathcal{D}} d\mathbf{r}_j \left(\frac{1}{r_i} + \frac{1}{r_j}\right)\delta(\theta_i - \theta_j + \pi - \epsilon)$$

$$= \frac{1}{4V^2} \int_0^{2\pi} d\theta_i\, r(\theta_i)r(\theta_i + \pi)[r(\theta_i) + r(\theta_i + \pi)] - \epsilon_0(\rho) \int_0^{2\pi} d\theta_i\, r(\theta_i)r'(\theta_i + \pi)[r(\theta_i) + 2r(\theta_i + \pi)]$$

$$- \epsilon_0^2(\rho) \int_0^{2\pi} d\theta_i\, r(\theta_i)r''(\theta_i + \pi)[r(\theta_i) + 2r(\theta_i + \pi)]. \tag{25}$$

In the polar coordinate centered at κ, the frontier of the disk is given by $r(\theta) = \sqrt{R^2 - \kappa^2 \sin^2(\theta)} - \kappa \cos(\theta)$ and the previous expressions can be now rewritten as

$$\overline{g(\kappa)} = \frac{1}{4V^2}\left[2\int_0^{2\pi} d\theta\, (R^2 - \kappa^2)\sqrt{R^2 - \kappa^2 \sin^2(\theta)} - \epsilon_0^2(\rho)\int_0^{2\pi} d\theta\, \frac{\kappa^2(R^2 - \kappa^2)(R^2(6\sin^2\theta - 1) - 5\kappa^2\sin^4\theta)}{(R^2 - \kappa^2\sin^2\theta)^{3/2}}\right]. \tag{26}$$

The first order term (coefficient of ϵ_0) is equal to 0 and the first nontrivial term is of second order. We introduce the functions

$$I_0(\kappa, R) = 2\int_0^{2\pi} d\theta\, (R^2 - \kappa^2)\sqrt{R^2 - \kappa^2 \sin^2(\theta)} \tag{27}$$

and

$$I_2(\kappa, R) = \int_0^{2\pi} d\theta\, \frac{\kappa^2(R^2 - \kappa^2)\{R^2[6\sin^2(\theta) - 1] - 5\kappa^2\sin^4(\theta)\}}{[R^2 - \kappa^2\sin^2(\theta)]^{3/2}}, \tag{28}$$

TABLE I. Details of the fit $\epsilon_0(\rho) = A\rho^{-\beta}$ for each family of graphs.

Graph family	A	β	R^2
k-NN	1.3 ± 0.2 (for $k = 7$)	0.3 ± 0.1	0.98
RGG	1.2 ± 0.2 (for $r = 2/\sqrt{\rho}$)	0.4 ± 0.1	0.95
DT	1.7 ± 0.2	0.5 ± 0.1	0.98
GG	2.7 ± 0.2	0.6 ± 0.1	0.98

and the BC can be rewritten as

$$\overline{g(\kappa)} = \frac{1}{4V^2}\left[I_0(\kappa) - \epsilon_0^2(\rho)I_2(\kappa, R)\right]. \tag{29}$$

In the result of Eq. (29), the infinite density limit which corresponds to the first term is universal, i.e., independent from the graph structure. In contrast, the second term [term in $\epsilon_0(\rho)$] does depend on the graph and encodes the deviation of shortest paths from the straight line which varies from a graph to another. This implies that in general (and as expected) the BC of a graph at finite density is not universal and depends on the graph considered. In particular, the numerical result of Fig. 3 suggests a power law relation of the form $\epsilon_0(\rho) \simeq A\rho^{-\beta}$ with $\beta \simeq 0.5 \pm 0.1$ (see Table I). We observe that the value of the exponent β and the prefactor are not exactly the same for all graphs, implying different rates of convergence toward the dense regime.

We can express the integrals appearing in Eq. (29) using special functions and we get

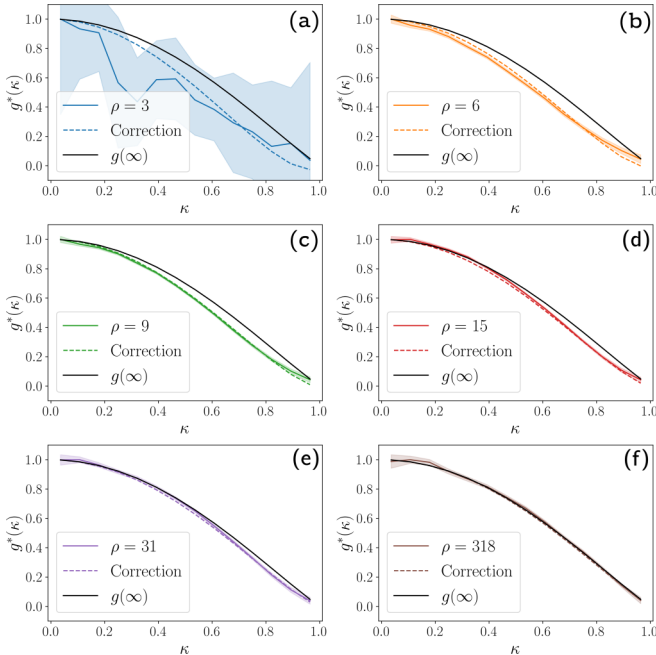$$I_0(R, \kappa) = 8R(R^2 - \kappa^2)E\left(\frac{\kappa}{R}\right) \tag{30}$$

FIG. 4. Comparison of the expansion of Eq. (35) with the numerical result for the DT. The quality of the approximation increases with the density but is insightful at surprisingly low densities (6 points per square unit is less than 20 points in the disk). The number of points is $N = \rho\pi$.
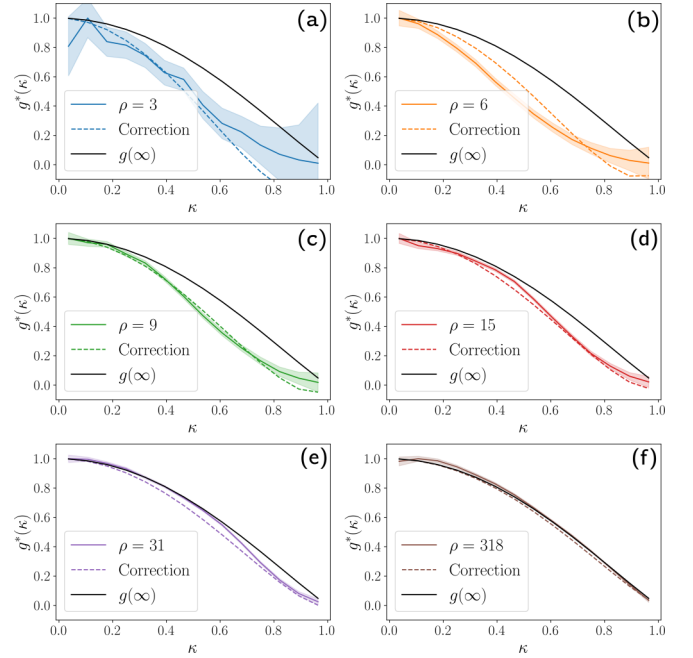

FIG. 5. Comparison of the expansion of Eq. (35) with the numerical result for the k-NN. The quality of the approximation increases with the density but it is already very good for a density $\rho = 9$ or larger. The number of points is $N = \rho\pi$.

and

$$I_2(R, \kappa) = 8R^3\left[3\left(\frac{\kappa}{R}\right)^2 K\left(\frac{\kappa}{R}\right) + 2K\left(\frac{\kappa}{R}\right) \tag{31}\right.$$

$$\left. + 2\left(\frac{\kappa}{R}\right)^2 E\left(\frac{\kappa}{R}\right) - 2E\left(\frac{\kappa}{R}\right)\right], \tag{32}$$

where $K(x)$ and $E(x)$ are, respectively, the elliptic integrals of the first and second kind.

If $\kappa \ll R$, then we can show that

$$I_2(R, \kappa) = 4\pi R^3\left(\frac{\kappa}{R}\right)^2 + o\left[\left(\frac{\kappa}{R}\right)^2\right], \tag{33}$$

while if $R - \kappa \ll R$, then

$$I_2(R, \kappa) \simeq 16R^3\left(1 - \frac{\kappa}{R}\right)\log\left(1 - \frac{\kappa}{R}\right). \tag{34}$$

Normalizing the BC by $\overline{g(0)} = \frac{1}{4V^2} \times 8\pi R^3$, we obtain

$$g^*(\kappa) = \frac{\overline{g(\kappa)}}{\overline{g(0)}} = \frac{1}{4\pi R^3}[I_0(\kappa) - \epsilon_0^2(\rho)I_2(\kappa, R)]. \tag{35}$$

If $\kappa \ll R$, then it gives

$$g^*(\kappa) \simeq 1 - \left[5 + \frac{1}{2}\epsilon_0^2(\rho)\right]\left(\frac{\kappa}{R}\right)^2, \tag{36}$$

while if $R - \kappa \ll R$, then

$$g^*(\kappa) \simeq \left[\frac{4}{\pi} - 2\epsilon_0^2(\rho)\log\left(1 - \frac{\kappa}{R}\right)\right]\left(1 - \frac{\kappa}{R}\right). \tag{37}$$

We note that under such a normalization we have always $g^*(0) = 1$ and $g^*(R) = 0$ (which can be easily proven). To use this result and to compare it with the simulations, we need to
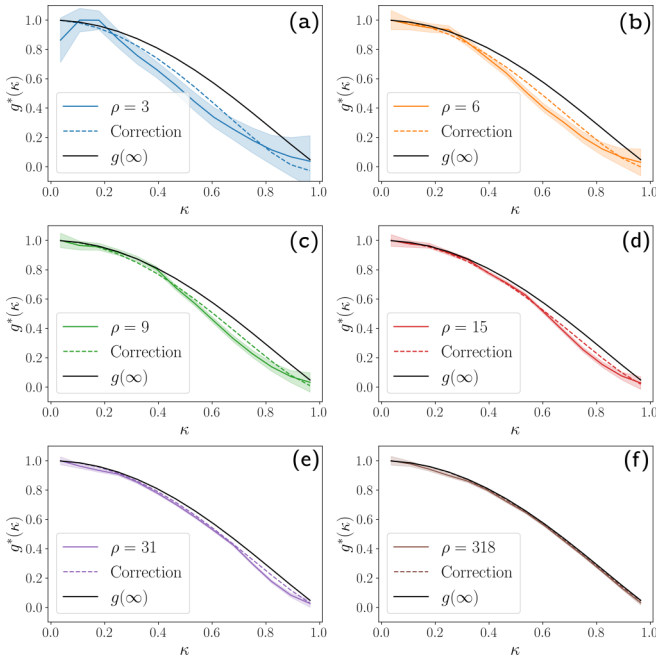
FIG. 6. Comparison of the expansion of Eq. (35) with the numerical result for the RGG. The convergence to the infinite density limit is faster than for other graphs and the approximation is very good for densities as low as $\rho = 9$. The number of points is $N = \rho\pi$.

specify the deviation characterized by $\epsilon_0(\rho)$ (see below for the numerical study).

C. Numerical study

We test the analytical result of Eq. (35) on various graphs: the DT (Fig. 4), the k-NN (Fig. 5), the RGG (Fig. 6), the GG (Fig. 7), and the MST (Fig. 8). For the numerical simulations, we sample N random vertices in a disk (we test sizes from $N = 10$ to $N = 1000$) and connect the points according to the rules of each graph. We then compute the BC of each point
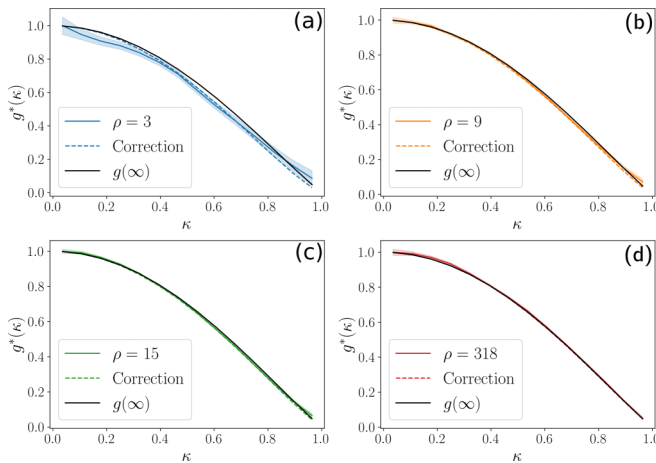

FIG. 7. Comparison of the expansion of Eq. (35) with the numerical result for the GG. The convergence to the infinite density limit is much faster than for other graphs and the approximation is very good for densities as low as $\rho = 3$. The number of points is $N = \rho\pi$.
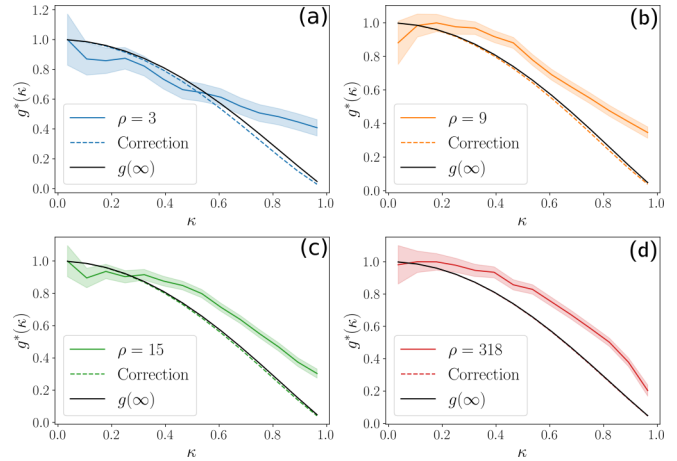

FIG. 8. Comparison of the expansion of Eq. (35) with the numerical result for the MST. The average BC converges toward the dense regime limit but the second-order approximation we used is not valid in this specific case.

using the Brandes algorithm [42] and average the results over 5000 runs (we build a new graph at each run).

For the GG, DT, k-NN, and the RGG, we have an excellent agreement between the analytical result and our numerical simulations (averaged over 5000 runs) for $1 \ll \rho$. The discrepancies between the quasidense and the dense regime are larger around $\frac{\kappa}{R} = 0.8$. We observe that for these different graphs the speed of convergence to the infinite density limit is not the same. The convergence for the GG, RGG, and the DT is fast while it is slower for the k-NN. In this respect, we show how the speed of convergence of k-NN graphs varies as a function of k on Fig. 9 and of the radius cutoff r for RGG on Fig. 10. For the GG, it is so fast that the infinite regime approximation is a good approximation for densities as low as 3 points per square unit (graphs of 10 points). In general, however, it is hard to predict or to understand why one type of graph would converge faster toward the infinite limit than
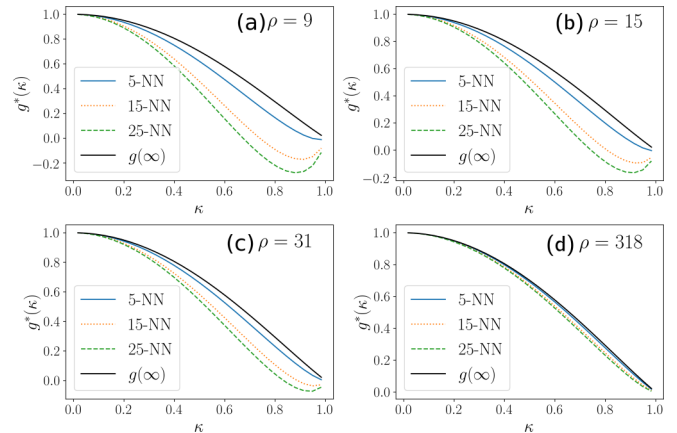

FIG. 9. Comparison of the speed of convergence of k-NN graphs toward the infinite density limit for different values of k: $k = 5$ (top line), $k = 15$ (middle line), and $k = 25$ (bottom line). The correction differs as the empirical fit of the prefactor A in Eq. (21) depends on the specific value of k.
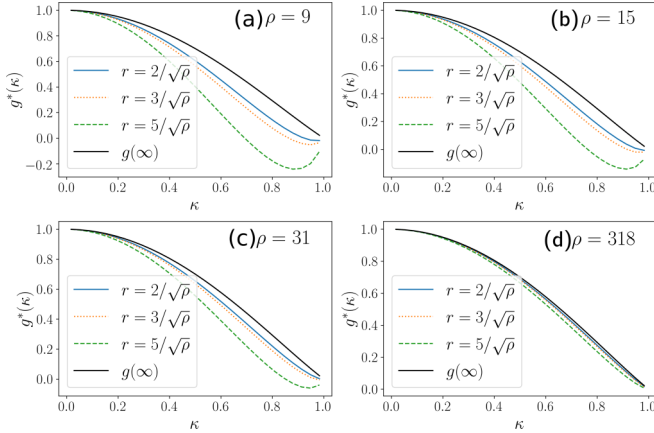
FIG. 10. Comparison of the speed of convergence of RGG graphs toward the infinite density limit for different values of the radius cutoff r.

FIG. 12. Comparison of the exact numerical result for the 2D lattice, the approximation Eq. (6), and the result obtained for the infinite density Eq. (30). We observe here that the infinite density result does not apply to this case.

the others. For k-NN and RGG graphs, the approximation is good for densities as low as 6 points per square unit (which corresponds to less than 20 points in the disk) while the approximation is valid for smaller densities (about 10 points in the disk) for DT and the Gabriel graphs (not shown here). We also note that Gabriel graphs (GG) being subgraphs of DT, we would naïvely expect that GG converge slower toward the dense regime limit than DT. This would result from the fact that the shortest paths are closer to straight lines (and hence the dense regime) when more points are added in the network. This is not true however since the normalized expected BC of κ depends on both the average BC in κ and the maximal BC in the graph (in 0). Adding more points to the system can both decrease the BC on average but increase the maximal BC, leading to nontrivial behaviors of convergence toward the

dense regime. We sum up the quality of the correction as a function of density for all families on graphs on Fig. 11.

Finally, we note that our approximation does not work for MST and RNG graphs since the assumption stating that $\chi(\epsilon)$ is independent of (i, j, κ) seems not to be valid for these graphs. The BC however converges toward the infinitely dense limit result of Ref. [33] (see Fig. 8 for the MST). At this point, it is an open question how to generalize our result to understand this behavior.

IV. NOTE: 2d GRID

We note that, somewhat surprisingly, for the 2D grid the above calculation doesn't apply. Indeed, we plot in Fig. 12 the exact numerical result for the 2D grid, the approximation Eq. (6), and the infinite density approximation [Eq. (30)]. There are two main reasons why the infinite density calculation does not apply here. First, there is a strong degeneracy and the number of shortest paths is very large in general, and second, these shortest paths are not straight. The main assumptions used to get the infinite density limit result [33] and our expansion do not therefore hold and we expect the observed discrepancy. However, the 2D grid case is not really a problem as we showed with the simple approximation Eq. (6).

V. DISCUSSION

In this paper we extended one of the few theoretical results about BC in spatial networks to a large number of families of graphs (k-NN, RGG, GG, and DT) for finite densities. We proved that for these families it is possible to find an approximation of the average BC of a random point in a bounded set of the plane as a function of its spatial coordinates. The infinite density limit which corresponds to the first term of our expansion is universal and independent from the graph. The first nontrivial correction encodes the deviations of shortest paths from the straight line and is therefore not universal. This approximation is theoretically valid for quasidense sets of points ($\rho \gg 1$) but is empirically correct for planar graphs with densities as low as a few points per square unit.
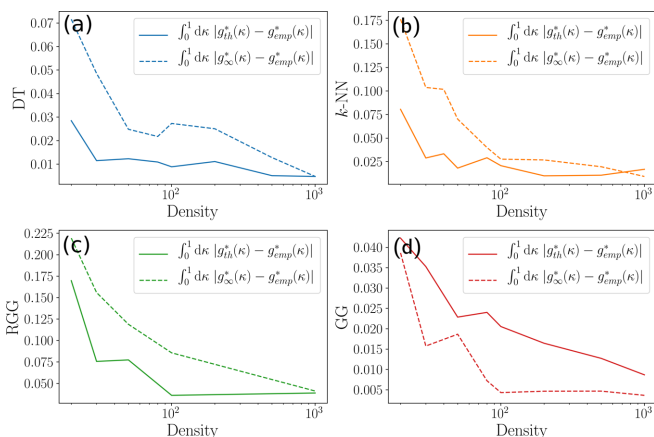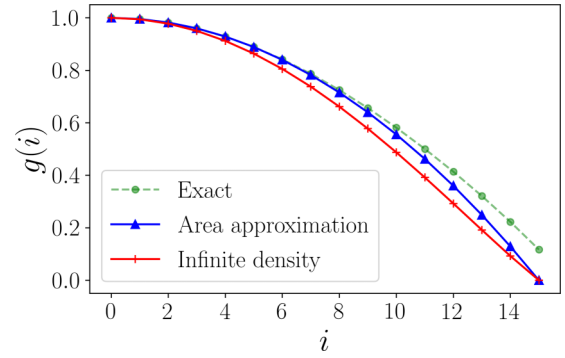

FIG. 11. Comparison of the quality of the correction as a function of point density for the different families of graphs. The quality is defined as the area between theoretical approximations (the infinite density limit g_∞^* and our expansion g_{th}^*) and the simulated (average) curves. The quality of both our approximation and the infinite density limit increases as the density does. As expected, the quality of the correction is better than the infinite density limit but the difference vanishes as the density increases.
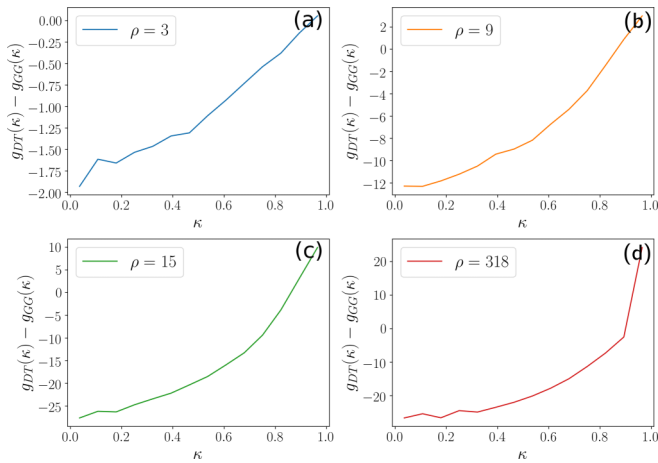
FIG. 13. On average, the BC decreases when constructing a Gabriel graph from a Delaunay triangulation. This results from the fact that the GG are subgraphs of DT. We note that the BC change is negative for all positions of κ except close to the edge of the disk, due to finite-size effects. These effects get more localized when increasing the density of points.

This approximation seems however not to be valid for other families of spatial networks (such as the RNG and the MST) whose dense limit is still universal but exhibit different convergence behaviors. The main difference comes from the way the shortest paths tend to straight lines and further studies are needed to understand this behavior.

We also observed that adding more points to the network decreases the BC on average (see Fig. 13) as theoretically expected. However, locally, some points may be more central when new points enter the system and it is therefore not possible to predict the speed of convergence toward the dense-regime from just an inclusion relation: if $G_1 \subset G_2$, there are more edges in G_2, but we can still have both a smaller average BC (due to theorem of Ref. [29]) and a larger maximal BC for this graph compared to G_1. For example, the Gabriel graph is a subgraph of the DT and we observe that the difference $g_{DT}(\kappa) - g_{GG}(\kappa)$ has a sign that can be either positive or negative according to the value of κ (Fig. 13), implying that the convergence to the infinite density limit is not "uniform."

The theoretical work proposed here is a further step toward the study the BC in spatial networks but many questions are still open. As we mentioned, it is unclear why the behavior of the MST and the RNG is so different from the other graphs studied here. More work is certainly needed to understand how shortest paths in these systems become always more straight when the density increases. Also, an open question concerns the spatial patterns of the BC in disordered spatial networks and it would interesting to understand from a theoretical point of view the effect of disorder.

[1] F. A. Rodrigues, Network centrality: An introduction, in *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems* (Springer, Cham, 2019), pp. 177–196.

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

[3] P. Holme, Congestion and centrality in traffic flow on complex networks, Adv. Complex Syst. **06**, 163 (2003).

[4] A. Jayasinghe, K. Sano, and H. Nishiuchi, Explaining traffic flow patterns using centrality measures, Int. J. Traffic Transport Eng. **5**, 134 (2015).

[5] A. Kazerani and S. Winter, Can betweenness centrality explain traffic flow, in *Proceedings of the 12th AGILE International Conference on Geographic Information Science* (Leibniz Universität Hannover, Germany, 2009), pp. 1–9.

[6] P. Santi and D. M. Blough, The critical transmitting range for connectivity in sparse wireless ad hoc networks, IEEE Trans. Mobile Comput. **2**, 25 (2003).

[7] J. Li, L. L. Andrew, C. H. Foh, M. Zukerman, and H. H. Chen, Connectivity, coverage, and placement in wireless sensor networks, Sensors **9**, 7664 (2009).

[8] J. Coon, C. P. Dettmann, and O. Georgiou, Impact of boundaries on fully connected random geometric networks, Phys. Rev. E **85**, 011138 (2012).

[9] I. Gupta, D. Riordan, and S. Sampalli, Cluster-head election using fuzzy logic for wireless sensor networks, in *Proceedings of the 3rd Annual Communication Networks and Services Research Conference (CNSR'05)* (IEEE, Piscataway, NJ, 2005), pp. 255–260.

[10] M. Ercsey-Ravasz and Z. Toroczkai, Centrality Scaling in Large Networks, Phys. Rev. Lett. **105**, 038701 (2010).

[11] O. Narayan and I. Saniee, Large-scale curvature of networks, Phys. Rev. E **84**, 066108 (2011).

[12] E. Jonckheere, M. Lou, F. Bonahon, and Y. Baryshnikov, Y. Euclidean versus hyperbolic congestion in idealized versus experimental networks, Internet Mathematics **7**, 1 (2011).

[13] S. Lämmer, B. Gehlsen, and D. Helbing, Scaling laws in the spatial structure of urban road networks, Physica A **363**, 89 (2006).

[14] P. Crucitti, V. Latora, and S. Porta, Centrality measures in spatial networks of urban streets, Phys. Rev. E **73**, 036125 (2006).

[15] S. Derrible, Network centrality of metro systems, PLoS ONE **7**, e40575 (2012).

[16] M. Barthelemy, P. Bordin, H. Berestycki, and M. Gribaudi, Self-organization versus top-down planning in the evolution of a city, Sci. Rep. **3**, 2153 (2013).

[17] E. Strano, V. Nicosia, V. Latora, S. Porta, and M. Barthelemy, Elementary processes governing the evolution of road networks, Sci. Rep. **2**, 296 (2012).

[18] M. Barthelemy, Betweenness centrality in large complex networks, Eur. Phys. J. B **38**, 163 (2004).

[19] A. Kirkley, H. Barbosa, M. Barthelemy, and G. Ghoshal, From the betweenness centrality in street networks to structural invariants in random planar graphs, Nat. Commun. **9**, 2501 (2018).

[20] Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley, Transport in Weighted Networks: Partition into Superhighways and Roads, Phys. Rev. Lett. **96**, 148702 (2006).

[21] S. Porta *et al.*, Street centrality and densities of retail and services in Bologna, Italy, Environ. Plan. B: Plan. Design **36**, 450 (2009).

[22] F. Wang, A. Antipova, and S. Porta, Street centrality and land use intensity in Baton Rouge, Louisiana, J. Transport Geogr. **19**, 285 (2011).

[23] S. Porta *et al.*, Street centrality and the location of economic activities in Barcelona, Urban Studies **49**, 1471 (2012).

[24] F. Wang, C. Chen, C. Xiu, and P. Zhang, Location analysis of retail stores in Changchun, China: A street centrality perspective, Cities **41**, 54 (2014).

[25] T. Davies and S. D. Johnson, Examining the relationship between road structure and burglary risk via quantitative network analysis, J. Quant. Criminol. **31**, 481 (2015).

[26] A. Venerandi, M. Zanella, O. Romice, J. Dibble, and S. Porta, Form and urban change–An urban morphometric study of five gentrified neighbourhoods in London, Environ. Plan. B: Urban Anal. City Sci. **44**, 1056 (2017).

[27] M. D. Scoppa and J. Peponis, Distributed attraction: The effects of street network connectivity upon the distribution of retail frontage in the City of Buenos Aires, Environ. Plan. B: Plan. Design **42**, 354 (2015).

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

[29] S. Gago, J. C. Hurajova, and T. Madaras, Notes on the betweenness centrality of a graph, Math. Slovaca **62**, 1 (2012).

[30] S. Gago, J. C. Hurajová, and T. Madaras, Betweenness centrality in graphs, *Quantitative Graph Theory: Mathematical Foundations and Applications, Discrete Mathematics and Its Applications* (Chapman and Hall/CRC, New York, 2014), pp. 233–257.

[31] B. Lion and M. Barthelemy, Central loops in random planar graphs, Phys. Rev. E **95**, 042310 (2017).

[32] A. Lampo, J. Borge-Holthoefer, S. Gomez, A. Solé-Ribalta, Multiple abrupt phase transitions in urban transport congestion, Phys. Rev. Research **3**, 013267 (2021).

[33] A. P. Giles, O. Georgiou, and C. P. Dettmann, Betweenness centrality in dense random geometric networks, in *Proceedings of the IEEE International Conference on Communications (ICC)* (IEEE, Piscataway, NJ, 2015), pp. 6450–6455.

[34] D. J. Aldous and J. Shun, Connected spatial networks over random points and a route-length statistic, Stat. Sci. **25**, 275 (2010).

[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] A. Auffinger, M. Damron, and J. Hanson, *50 Years of First-passage Percolation* (American Mathematical Society, Providence, RI, 2017), Vol. 68.

[37] A. P. Kartun-Giles, K. Koufos, and N. Privault, Connectivity of 1D random geometric graphs, arXiv:2105.07731.

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

[39] M. Walters, Random geometric graphs, Surveys Combinator. **392**, 365 (2011).

[40] M. Haenggi, *Stochastic Geometry for Wireless Networks* (Cambridge University Press, Cambridge, UK, 2012).

[41] L. A. Santalò, *Integral Geometry and Geometric Probability* (Cambridge University Press, Cambridge, UK, 2004).

[42] U. Brandes, A faster algorithm for betweenness centrality, J. Math. Sociol. **25**, 163 (2001).