Results in Mathematics

Continuous Mean Distance of a Weighted Graph

Delia Garijo, Alberto Márquez, and Rodrigo I. Silveira

Abstract. We study the concept of the *continuous mean distance* of a weighted graph. For connected unweighted graphs, the *mean distance* can be defined as the arithmetic mean of the distances between all pairs of vertices. This parameter provides a natural measure of the compactness of the graph, and has been intensively studied, together with several variants, including its version for weighted graphs. The *continuous* analog of the (discrete) mean distance is the mean of the distances between all pairs of *points* on the edges of the graph. Despite being a very natural generalization, to the best of our knowledge this concept has been barely studied, since the jump from discrete to continuous implies having to deal with an infinite number of distances, something that increases the difficulty of the parameter. In this paper, we show that the continuous mean distance of a weighted graph can be computed in time roughly quadratic in the number of edges, by two different methods that apply fundamental concepts in discrete algorithms and computational geometry. We also present structural results that allow for a faster computation of this continuous parameter for several classes of weighted graphs. Finally, we study the relation between the (discrete) mean distance and its continuous counterpart, mainly focusing on the relevant question of convergence when iteratively subdividing the edges of the weighted graph.

Mathematics Subject Classification. 65D18, 68R10.

Keywords. Weighted graph, Mean distance, Wiener index, Geometric graph, Algorithms.

1. Introduction

Distances are one of the most essential aspects in the analysis of graphs, regardless of whether they originate in geography, transportation, sociology, or

communications. The maximum distance between any two nodes in the graph, known as the *diameter*, provides a worst-case scenario in terms of distances, and gives the maximum *eccentricity* in the graph. Similarly, the *average* or *mean distance* is related to centrality, and provides a measure of the compactness of the graph. In this work we will focus on the latter, the mean distance.

The *mean distance* of a connected unweighted graph $G = (V(G), E(G))$ was first introduced by March and Steadman [19, Chap.14] in the context of architecture to compare floor plans, although interest in the concept dates back to the work of Wiener in chemistry [35] (after whom the closely related *Wiener index*, the sum of all pairwise distances in the graph, is named).

The most usual way to define the mean distance $\mu(G)$ is as the arithmetic mean of all nonzero distances between vertices, i.e.,

$$\mu(G) = \frac{1}{\binom{|V(G)|}{2}} \sum_{\{u,v\} \subseteq V(G)} d(u,v), \tag{1}$$

where $|V(G)| \geq 2$, $d(u,v)$ is the length of a shortest path connecting vertices u and v, and the sum is taken over all unordered pairs of vertices in the graph.

In the context of graph theory, Doyle and Graver [8] were the first to propose the mean distance as a graph parameter. Since them, it has been intensively studied. For very simple graphs, the mean distance is well-understood. For instance, it is 1 in any complete graph, and $(n+1)/3$ if the graph is an n-vertex path. However, as soon as the graph becomes more complicated, the expression for its mean distance becomes much more elusive. In addition to presenting exact expressions for a few specific graph classes [2,3,8], most previous work has focused on proving lower and upper bounds on the mean distance as a function of parameters such as the number of vertices [8,11,27], number of vertices and edges [32], and connectivity [12]. Considerable effort was also put into understanding the relation between the mean distance and the minimum vertex degree [17], as well as some spectral graph properties [20,21,30].

The concept of mean distance has also been extended to weighted graphs, both for vertex weights [6], and for rational edge weights [9,10]. A few have also studied the concept for directed graphs [22,27].

As mentioned above, a concept that is closely related to the mean distance is the Wiener index, defined as the sum of distances between all (unordered) pairs of vertices in the graph. The Wiener index has been studied extensively (both for unweighted and weighted graphs) due to its important applications in chemical graph theory [23], but has also received attention in other areas, such as mathematics [16] and social graph analysis [25], and it is still the topic of active investigation (see, e.g., [31]). From the point of view of computation, there have been important efforts in computer science to understand how efficiently the Wiener index can be computed. While it is immediate to obtain a roughly quadratic-time algorithm that computes each distance in the graph (e.g., by solving an all-pairs shortest path problem), the challenge is to understand in which situations this can be done more efficiently. Since for arbitrary graphs it

is known that this is not possible unless the strong exponential time hypothesis (SETH) fails [29], the focus has been on identifying classes of graphs for which the Wiener index can be computed in subquadratic time. Some cases for which this has been shown to be possible is for graphs with bounded treewidth [1,5], and most notably for planar graphs with non-negative cycles [4].

Going back to the mean distance, a different direction was adopted by Doyle and Graver [9,10], who introduced the mean distance of a *shape.* This is defined for any weighted graph embedded in the plane. Each edge of the graph is iteratively subdivided into shorter edges, so that the edge lengths approach zero. The mean distance of the shape is then defined as the limit of the mean distance of such a sequence of refinements. While this is a natural definition, its computation is involved. Doyle and Graver managed to compute its exact value for seven specific types of simple graphs (i.e., a path, a Y-shape, an H-shape, a cross, and three more) and six rather specific families of graphs; the most general ones being cycles and stars with k edges of length $1/k$. Examples of the more specific families studied are graphs consisting of one edge with two edges attached at each endpoint, and (multi)graphs consisting of three edges sharing both endpoints, in both cases for very constrained edge lengths. A summary of these formulas is given in [10]; they are obtained as a consequence of the techniques developed in [9], mainly, for trees and the so-called *geometric* shapes.

In this paper, we study the mean distance in a *continuous* setting, in a spirit very similar to that of the shapes of Doyle and Graver [9,10]. Our main motivation arises from *geometric graphs.* A geometric graph is an undirected graph where each vertex is a two-dimensional point, and each edge is a straight line segment between the corresponding two points. Geometric graphs appear naturally in many applications, for instance in road, river or computer graphs. Unlike abstract graphs, in geometric graphs distances are not only defined for pairs of vertices, but they exist for any two points on the graph, including points on the interior of edges. Therefore, the concept of mean distance generalizes naturally to (weighted) geometric graphs, defined as the average distance between all *pairs of points* on edges of the graph. While being a natural definition, the jump from discrete to continuous implies that now the mean is the sum of an infinite number of distances, something that changes the properties of this index and makes its computation difficult. In this paper, we study this concept in depth, with the focus on the computational aspects of the continuous mean distance, and on understanding how much it differs from the vastly studied discrete mean distance.

In particular, our main contributions are:

- We show that the continuous mean distance of a weigthed graph with m edges can be computed in $O(m^2)$ time, once all pair-wise distances between vertices have been computed. To this end, we present two different methods, one based on a generalization of shortest path trees to

continuous distances, and one based on Voronoi diagrams for the L_1 (or Manhattan) metric. See Sect. 3.

- We present several structural results that allow a faster computation of the continuous mean distance for several classes of weighted graphs. In particular, we give an exact expression for complete graphs where all edges have the same length, and efficient algorithms for families of graphs that have a cut vertex, which include weighted trees and weighted cactus graphs. See Sect. 4.

- We study the relation between the discrete mean distance and the continuous counterpart. After establishing some relations between them in Sect. 5, we move to the relevant question of convergence: When does iteratively subdividing edges and computing the discrete mean distance converge to the continuous mean distance? While a definitive answer to this question does not seem possible, in Sect. 6 we study a refining procedure that gives a guarantee on how much the discrete and continuous means can differ as the weighted graph is iteratively refined. The bounds obtained are tight for some graphs classes, such as trees where all edges have the same length.

Next we present our problem formally.

2. Preliminaries

Let $G = (V(G), E(G))$ be a connected graph[1] with n vertices and m edges; when no confusion may arise, we indistinctly write V or $V(G)$ and E or $E(G)$. Consider a function $\omega : E \longrightarrow \mathbb{R}^+$ that assigns a positive weight $\omega(e)$ to each edge $e \in E$. The value $\omega(e)$ is called the *length* of edge e, and is also denoted by $|e|$. In general, given a subset of edges $E' \subseteq E$, its *weight* or *length* is $|E'| = \sum_{e \in E'} \omega(e)$.

Graph G together with function ω is a weighted graph where every edge can be identified with a line segment of length $\omega(e)$ in the Euclidean plane. Thus, every point p on an edge $e = uv$ can be expressed as $p = \lambda_p v + (1 - \lambda_p)u$ for some $\lambda_p \in [0, 1]$. Let G_ℓ be the set of all points that are on the edges on G. Note that this definition not only includes all geometric graphs, but also covers other graphs that are not geometric. A simple example of such a graph is a triangle where two edges have length 1 and the third one has length 2; such a graph cannot be realized with three straight line segments, since it would require the longer edge to overlap with the two shorter ones.

We point out that all graphs considered in this work are connected and weighted, although both terms will be in general omitted as it is understood from the context. We will also consider *uniform* graphs: graphs where all edges have the same length. We will write α-*uniform* to refer to a uniform graph where all edge lengths are α.

---

[1] All graphs considered in this work are assumed to be connected.

Let p, q be two points on G_ℓ that are not both on the interior of the same edge. A *path* P_ℓ between p and q, also called *pq-path*, is a sequence $p u_1 \ldots u_k q$ where and the *distance* $d(p, q)$ between p and q on G_ℓ is the length of a shortest path connecting the two points. When the two points p, q are on the interior of the same edge $u_0 u_1$ and, say $\lambda_p < \lambda_q$, we have paths between p and q that go through vertices (whose definition is analogous to the above one) but also a path in the interior of the edge that is the segment connecting p and q (edges are identified with segments), and its length is $(\lambda_q - \lambda_p)\omega(u_0 u_1)$. In this paper, we shall assume that the distance between the two endpoints of any edge e is $|e|$. The set of points G_ℓ together with this distance function is a metric space, and it will be treated indistinctly as a graph (with vertex set $V(G_\ell) = V(G)$ and edge set $E(G_\ell) = E(G)$) or as a point set. The distance between an edge $e = uv$ and a point $p \notin e$ is $d(p, e) = \min\{d(p, u), d(p, v)\}$ (if $p \in e$, $d(p, e) = 0$), and the distance between two edges e and $e' = ab$ is $d(e, e') = \min\{d(a, e), d(b, e)\}$.

We begin by defining the variant of the *discrete mean distance* that we will consider in the remainder of this work. The definition below differs from the Eq. (1) of $\mu(G)$ in two aspects: (i) it considers *all* pairs of distances, including those that are zero, and (ii) it considers *ordered* pairs of vertices:

$$\mu_d(G) = \frac{1}{n^2} \sum_{(u,v) \in V \times V} d(u, v) = \frac{2W(G)}{n^2}, \qquad (2)$$

where $W(G)$ denotes the Wiener index of G. Observe that $\mu_d(G)$ is the arithmetic mean of the entries of the distance matrix of the graph. Although this alternative form of mean distance has been considered before [34], our motivation for studying it comes from the fact that it extends better to the continuous mean distance (which is the subject of this paper) in a limiting process when iteratively subdividing the edges of the graph. In particular, it will allow us to establish a clear relation between the discrete and the continuous mean distance.

To define formally the continuous mean distance of a weighted graph, we start by defining it between a point and a set of edges. Given a point $p \in G_\ell$ and a subset of edges $E' \subseteq E(G_\ell)$, the *continuous mean distance* between p and E' is

$$\mu_c(p, E') = \frac{1}{|E'|} \int_{q \in E'} d(p, q) \, dq. \qquad (3)$$

For subsets of edges $E', E'' \subseteq E(G_\ell)$, the *continuous mean distance* between E' and E'' is

$$\mu_c(E', E'') = \frac{1}{|E'||E''|} \iint_{p \in E', \, q \in E''} d(p, q) \, dp \, dq. \qquad (4)$$

With some abuse of notation, we shall write $\mu_c(p, G')$ or $\mu_c(G', G'')$, where G' and G'' are the graphs with edge sets E' and E'', respectively.

Based on the previous, the *continuous mean distance* of the weighted graph G_ℓ is defined as

$$\mu_c(G_\ell) = \mu_c(E(G_\ell), E(G_\ell)). \tag{5}$$

Remark 2.1. In [9, Equations (1) and (3)], the authors show that the discrete mean distance of an α-uniform graph G can be easily deduced from the 1-uniform case. They use definition (1) and $\alpha \in \mathbb{Q}^+$: $\mu(G) = \alpha\mu(G_1)$ where G_1 is the corresponding 1-uniform graph. This can be naturally extended to the variant $\mu_d(G)$ and $\alpha \in \mathbb{R}^+$, and by elementary properties of integration, a similar formula holds for the continuous mean distance even when the graph G_ℓ is not uniform: $\mu_c(G'_\ell) = \beta\mu_c(G_\ell)$ where G'_ℓ is the graph obtained by simply multiplying all edge lengths of G_ℓ by β.

The following observations, which follow directly from Eq. (4), will be used throughout this work.

Remark 2.2. Let ab and uv be two edges in $E(G_\ell)$.

(i) For any point $p \in uv$ we have

$$\mu_c(uv, ab) = \frac{|up|\mu_c(up, ab) + |pv|\mu_c(pv, ab)}{|uv|}.$$

(ii) For each point $p \in uv$ and each point $q \in ab$, if $d(p,q) = d(p,v) + d(v,q)$, we have

$$\mu_c(uv, ab) = \mu_c(uv, v) + \mu_c(v, ab).$$

An example: paths It is illustrative to see how the continuous mean distance can differ from the discrete version. Here we illustrate this for the important case of paths.

Consider a 1-uniform path P, i.e., a graph consisting of a path with n vertices and all edges of length 1. The discrete mean distance of such a path is known to be $\mu_d(P) = (n^2 - 1)/3n$ [34]. By Remark 2.1, this generalizes to $\mu_d(P) = \alpha(n^2 - 1)/3n$ when P is α-uniform (its total length is $\alpha(n - 1)$). For non-uniform paths P, there is no closed formula to compute $\mu_d(P)$. In contrast, it is possible to obtain a closed formula for $\mu_c(P_\ell)$, for any path with arbitrary positive real edge lengths, as explained next.

First observe that, for the continuous mean distance, the number of interior nodes in a path does not play any role, thus we can consider the path as one single edge. Hence a path P_ℓ of length $t \in \mathbb{R}_{>0}$ can be seen as the interval $[0, t]$. For a point $x \in [0, t]$, let $d(x, [0, t])$ denote the function that gives the distance between x and any other point x' in the interval $[0, t]$; the shape of this function is illustrated in Fig. 1. The mean value of $d(x, [0, t])$ is $\frac{1}{2t}(x^2 + (x - t)^2)$.[2] Thus,

$$\mu_c(P_\ell) = \frac{1}{t} \int_0^t \frac{1}{2t}(x^2 + (x - t)^2)\, dx = \frac{t}{3}. \tag{6}$$

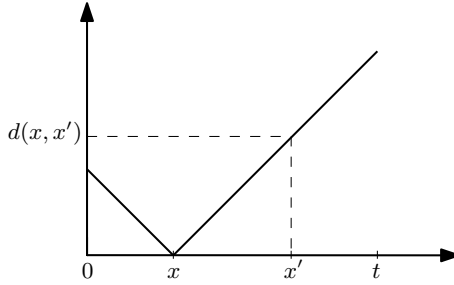[2]Recall that the mean value of a function f over an interval $[a, b]$ is $\frac{1}{b-a} \int_a^b f(x)\, dx$.

FIGURE 1. Function $d(x, [0, t])$

Based on a different approach, the same value was given in [9] (see also [10]) for α-uniform paths P_ℓ with $\alpha \in \mathbb{Q}^+$.

3. Computation of the Continuous Mean Distance

The continuous nature of the continuous mean distance makes its computation non-trivial, as exemplified by the seemingly simple case of paths. In this section, we show that despite this, $\mu_c(G_\ell)$ can be computed rather efficiently, in time roughly quadratic in the number of edges of G_ℓ. We will show how this can be achieved in two different ways, which apply some fundamental concepts in discrete algorithms and computational geometry: that of shortest path trees and that of Voronoi diagrams for the L_1 (or Manhattan) metric. We highlight that the relation between the continuous mean distance and these two ubiquitous structures is interesting on its own.

The main result of this section is the following.

Theorem 3.1. *The continuous mean distance of a weighted graph G_ℓ with n vertices and m edges can be computed in $O(m^2 + A(n, m))$ time, where $A(n, m)$ is the time required to compute all vertex-to-vertex distances in G_ℓ.*

To prove the preceding theorem, we use the following formula, which states that $\mu_c(G_\ell)$ can be obtained as a weighted sum of the continuous mean distances of all *ordered* pairs of edges; this is simply a consequence of Eqs. (4)–(6), and elementary properties of integration.

$$\mu_c(G_\ell) = \frac{1}{|E|^2} \left(\sum_{(e,e') \in E \times E, e \neq e'} \mu_c(e, e')|e||e'| + \sum_{e \in E} \frac{|e|}{3}|e|^2 \right). \tag{7}$$

This fact reflects that understanding how the continuous mean distance behaves in the case of two edges is the key tool to compute it for the whole graph. In the next subsections, we present our two different approaches for the two-edge case. Theorem 3.2 as well as Theorem 3.3 let us conclude that the continuous mean distance between two edges can be computed in constant
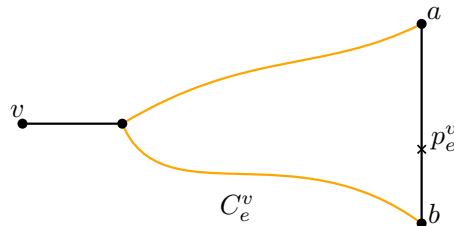
FIGURE 2. Point p_e^v, for $e = ab$, is the point furthest from v on cycle C_e^v (composed of the orange paths and edge ab)

time, once the distance matrix of the vertices of the graph G has been computed. The currently best algorithm to compute all-pairs shortest paths for a graph with real weights has running time $A(n, m) = O(nm \log \alpha(m, n))$ [26], where $\alpha(m, n)$ is the extremely slowly growing inverse of the Ackermann function. We note that for some special graph classes faster algorithms are known, such as planar graphs with non-negative edge weights (where $A(m, n) = O(n^2)$ [14]), or graphs with integer non-negative edge weights (for which $A = O(nm)$ [33]).

The continuous mean distance of two equal edges reduces to the mean distance of a path, which—as we have seen in Eq. (6)—is equal to $\mu_c(e, e) = |e|/3$, for any edge e; this is used in Eq. (7). Therefore, in the remainder of the section, we focus on the mean distance between two distinct edges.

3.1. Computation Using Shortest Path Trees

Shortest path trees are one of the most fundamental structures used to represent distances in graphs, and they are an essential underlying concept behind most single-source shortest path algorithms. In this section, we introduce a continuous version of the shortest path tree rooted at a vertex of a weighted graph G, and later show how it can be used to compute the continuous mean distance between any two distinct edges of G_ℓ.

For G_ℓ and a vertex $v \in V(G_\ell)$, a *continuous shortest path tree* is a pair $\mathcal{T}_v = (T_v, S_v)$, where T_v is a (discrete) shortest path tree rooted at v, and S_v is a subset of G_ℓ that contains one point p_e^v for each edge $e \in E(G_\ell) \backslash E(T_v)$. Point p_e^v is the only point on edge $e = ab$ such that its distance to v is given by two different paths: one passes through a, and the other one passes through b. Thus, p_e^v is the furthest point to v on any cycle C_e^v determined by e and shortest paths connecting v with a and b (see Fig. 2). Note that p_e^v must exist, otherwise $e \in E(T_v)$. Observe also that for any point on ap_e^v, its shortest paths to v go through a, and analogously for points on $p_e^v b$.

As we show next, the continuous shortest path tree can be computed within the same running time needed to solve the single-source shortest path problem, denoted by $S(n, m)$. Currently, we have $S(n, m) = O(m \log \alpha(m, n))$

time in general [26], in $S(n,m) = O(n)$ time for planar graphs with non-negative edge weights [14], and in $S(n,m) = O(m)$ time for graphs with integer non-negative edge weights [33].

Proposition 3.1. *Let $G = (V, E)$ be a weighted graph with n vertices and m edges, and let $v \in V$. A continuous shortest path tree $\mathcal{T}_v = (T_v, S_v)$ of G_ℓ can be computed in $O(S(n,m))$ time, where $S(n,m)$ is the time required to compute a shortest path tree from v.*

Proof. We first compute T_v using a single-source shortest path algorithm in $O(S(n,m))$ time. Now, let $e = ab \in E \backslash E(T_v)$ and $p_e^v \in S_v$. Since $p_e^v = \lambda_e b + (1 - \lambda_e)a$ for some $\lambda_e \in [0,1]$, and p_e^v is the furthest point from v on any cycle C_e^v, we have:

$$\lambda_e |ab| + d(v,a) = d(v,p_e^v) = \frac{|ab| + d(v,b) + d(v,a)}{2} \Rightarrow \lambda_e$$
$$= \frac{|ab| + d(v,b) - d(v,a)}{2|ab|}.$$

Thus, λ_e can be computed in constant time. Therefore, the set S_v can be computed in total $O(m)$ time, once the single-source shortest path tree is available. $\square$

Depending on whether an edge e belongs to $E(T_v)$, Lemma 3.1 below provides a different expression for its continuous mean distance to vertex v. In the proof of this lemma, and throughout this paper, we shall use that, by the mean value theorem for integrals, the mean value of a function f over an interval $[a, b]$ coincides with the height of the rectangle with base $b - a$ and area $\int_a^b f(x)\,dx$.

Lemma 3.1. *Let $v \in V(G_\ell)$ and $e = ab \in E(G_\ell)$. Then,*

$$\mu_c(v,e) = \begin{cases} \min\{d(v,a), d(v,b)\} + \dfrac{|ab|}{2} & \text{if } e \in E(T_v), \\ \left(d(v,a) + \dfrac{|ap_e^v|}{2}\right)\lambda_e^v + \left(d(v,b) + \dfrac{|p_e^v b|}{2}\right)(1 - \lambda_e^v) & \text{if } e \notin E(T_v), \end{cases}$$

where $|ap_e^v| = \dfrac{|ab| + d(v,b) - d(v,a)}{2}$ and $|p_e^v b| = \dfrac{|ab| + d(v,a) - d(v,b)}{2}$.

Proof. Suppose first that $e \in E(T_v)$. Figure 3a illustrates the graph of $d(x,v)$ for $x \in e$ assuming that $d(a,v) < d(b,v)$ (analogous otherwise). This is a straight-line segment with slope 1, so the height of the rectangle with base $b - a$ and area $\int_a^b d(x,v)\,dx$ is $d(a,v) + |ab|/2$. Hence, the result follows.

Assume now that $e \notin E(T_v)$. We can argue as above but considering two rectangles determined by the function $d(x,v)$, one for $x \in [a, p_e^v]$ and the other for $[p_e^v, b]$, where $p_e^v = \lambda_e b + (1 - \lambda_e)a$ for some $\lambda_e \in [0,1]$; see Fig. 3b. The heights of these rectangles are, respectively, $d(v,a) + \frac{1}{2}|ap_e^v|$ and $d(v,b) + \frac{1}{2}|p_e^v b|$.
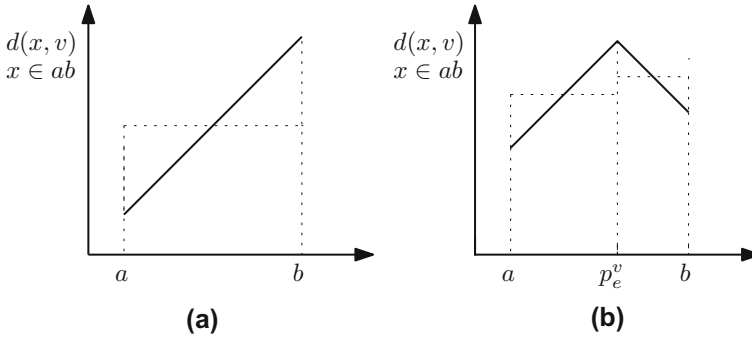
FIGURE 3. Function $y = d(x, v)$ when: **a** $e \in E(T_v)$, **b** $e \notin E(T_v)$

To obtain $\mu_c(v, e)$ each of these values must be multiplied by the proportion of the segment that corresponds to the base.

The expressions for $|ap_e^v|$ and $|p_e^v b|$ come from the fact that p_e^v is the farthest point from v on any cycle C_e^v (see Fig. 2). Thus,

$$d(v, a) + |ap_e^v| = \frac{|ab| + d(v, b) + d(v, a)}{2} = d(v, b) + |bp_e^v|.$$

$\square$

The preceding lemma will be used to compute the continuous mean distance between two edges. To do this, we distinguish three cases (Lemmas 3.2–3.4 below), two of which depend on the following property.

Property 3.1 (Same component property). *Let $ab \in E(G_\ell)$. A vertex $v \in V(G_\ell) \backslash \{a, b\}$ satisfies the* same component property *with respect to edge ab and vertex a if the shortest path from a to v goes through b.*

The same component property essentially means that all shortest paths from points on edge ab to v go through b.

Lemma 3.2 (Rectangular case). *Let $ab, uv \in E(G_\ell)$ be two distinct edges such that:*

(i) $|ab| = |uv| = \lambda$,

(ii) $d(a, uv)$ *is given by an au-path of length θ, and $d(b, uv)$ is given by a bv-path of the same length,*

(iii) *the paths of (ii) do not intersect.*

Then $\mu_c(ab, uv) = \theta + \dfrac{2\lambda}{3}$.

Proof. Edges ab and uv can be seen as the interval $[0, \lambda]$. By Eq. (4), we have:

$$\mu_c(ab, uv) = \frac{1}{\lambda^2} \int_0^\lambda \int_0^\lambda d(x, y) \, dx \, dy = \frac{1}{\lambda} \int_0^\lambda \mu_c(x, uv) \, dx$$
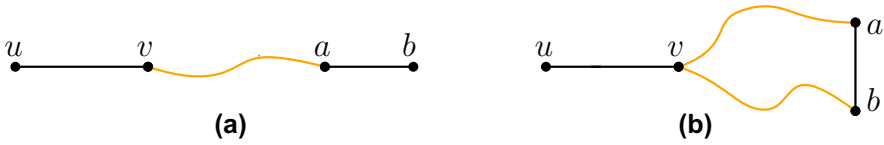
FIGURE 4. Linear cases: **a** $ab \in E(T_u)$, **b** $ab \notin E(T_u)$

Given a point $x \in [0, \lambda]$ on edge ab, its furthest point on the cycle composed by ab, uv, the au-path of length θ, and the bv-path of the same length is the point $\lambda - x$ on edge uv. This point plays the role of p_e^v in Lemma 3.1 with $\lambda_e^v = (\lambda - x)/\lambda$. Note that Lemma 3.1 is stated for vertices and edges of the graph, but we can always insert a vertex at the required point x and consider, with some abuse of notation, the tree T_x. Since the shortest $x - u$ and $x - v$ paths do not contain the edge uv, we have $uv \notin E(T_x)$. Then, by Lemma 3.1,

$$\mu_c(x, uv) = \left(x + \theta + \frac{\lambda - x}{2} \right) \frac{\lambda - x}{\lambda} + \left(\lambda - x + \theta + \frac{x}{2} \right) \frac{x}{\lambda},$$

where $d(x, u) = x + \theta$ and $d(x, v) = \lambda - x + \theta$. Hence,

$$\mu_c(ab, uv) = \frac{1}{\lambda^2} \int_0^\lambda \left[\left(x + \theta + \frac{\lambda - x}{2} \right) (\lambda - x) + \left(\lambda - x + \theta + \frac{x}{2} \right) x \right] dx.$$

Thus,

$$\mu_c(ab, uv) = \frac{1}{\lambda^2} \int_0^\lambda \left(-x^2 + \lambda x + \lambda \theta + \frac{\lambda^2}{2} \right) dx = \theta + \frac{2\lambda}{3}.$$

$\square$

Lemma 3.3 (Linear case). *Let $ab \in E(G_\ell)$ be an edge such that a and b satisfy the same component property with respect to other edge uv and one of its endpoints, say u. Then,*

$$\mu_c(uv, ab) = \frac{|uv|}{2} + \mu_c(v, ab).$$

Proof. In this case, there are two possible situations that may happen for ab and uv, see Fig. 4. For each $p \in uv$ and each $q \in ab$, we have $d(p, q) = d(p, v) + d(v, q)$. By Remark 2.2(ii), it follows that $\mu_c(uv, ab) = \mu_c(uv, v) + \mu_c(v, ab) = \frac{|uv|}{2} + \mu_c(v, ab)$. $\square$

Lemma 3.4 (Cycle case). *Let $ab, uv \in E(G_\ell)$ be two distinct edges such that neither u, v nor a, b satisfy the same component property with respect to the other corresponding edge and one of its endpoints. Then, $\mu_c(ab, uv)$ can be computed as a weighted sum of at most four linear cases and one rectangular case.*
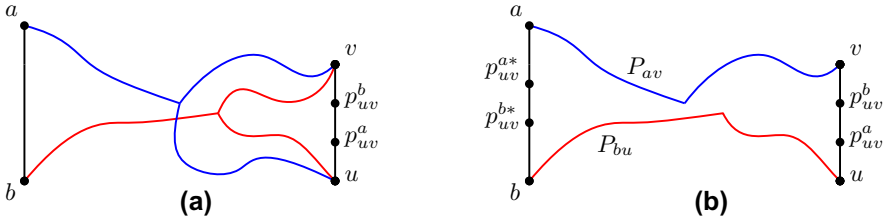
FIGURE 5. **a** Trees T_a (in blue) and T_b (in red), and points p_{uv}^a and p_{uv}^b. **b** Cycle $\mathcal{C}$ formed by the two edges and the paths P_{av} and P_{bu}; the points p_{uv}^{a*} and p_{uv}^{b*} are the furthest points of, respectively, p_{uv}^a and p_{uv}^b on $\mathcal{C}$ (color figure online)

Proof. Consider the continuous shortest path trees (T_a, S_a) and (T_b, S_b) rooted at a and b, respectively. Figure 5a illustrates how T_a and T_b must be located with respect to edges ab and uv, since no endpoint of these two edges satisfy the same component property. Note that T_a and T_b might have paths in common. Observe also that $uv \notin E(T_a) \cap E(T_b)$.

Let $p_{uv}^a \in S_a$ and $p_{uv}^b \in S_b$. Suppose first that the two points are distinct and, assume without loss of generality, that p_{uv}^a is closer to u than p_{uv}^b, see Fig. 5a. Applying Remark 2.2(i) twice, we have $\mu_c(ab, uv) = \frac{|up_{uv}^a|}{|uv|}\mu_c(ab, up_{uv}^a) + \frac{|p_{uv}^a p_{uv}^b|}{|uv|}\mu_c(ab, p_{uv}^a p_{uv}^b) + \frac{|p_{uv}^b v|}{|uv|}\mu_c(ab, p_{uv}^b v)$. Further, $\mu_c(ab, up_{uv}^a)$ can be computed by Lemma 3.3 as a linear case: it suffices to consider the shortest paths in T_a and T_b connecting, respectively, a and b with u; vertices a and b satisfy the same component property with respect to $p_{uv}^a u$ and point p_{uv}^a. The situation is analogous for $\mu_c(ab, p_{uv}^b v)$, and it remains to obtain $\mu_c(ab, p_{uv}^a p_{uv}^b)$. Note that Property 3.1 is stated for vertices and edges of the graph, but one can always insert vertices at the required points, such as p_{uv}^a, in order to deal with the situation as a linear case.

We now consider the cycle $\mathcal{C}$ determined by edges ab and uv, and the shortest paths in T_a and T_b giving, respectively, $d(a, uv)$ and $d(b, uv)$. Suppose, without loss of generality that those paths, denoted by P_{av} and P_{bu}, connect a with v and b with u (see Fig. 5b). Observe that $\mathcal{C}$ is indeed a cycle as $p_{uv}^a \neq p_{uv}^b$; otherwise we would have a common sub-path in P_{av} and P_{bu}, so there would be a point satisfying that its further point on uv would be the same as the furthest points of a and b on uv, which would imply $p_{uv}^a = p_{uv}^b$. Note also that C is a cycle of minimum length containing edges ab and uv since $|P_{av}| = d(a, uv)$ and $|P_{bu}| = d(b, uv)$.

Let p_{uv}^{a*} be the point furthest from p_{uv}^a on $\mathcal{C}$, and let p_{uv}^{b*} be defined analogously for p_{uv}^b. Refer to Fig. 5b. By construction of the cycle, these points are on edge ab, since $\mathcal{C}$ contains a shortest ap_{uv}^a-path and a shortest bp_{uv}^b-path. This implies that the furthest point of a on $\mathcal{C}$ is either p_{uv}^a or is located in between p_{uv}^a and b. Analogously, the furthest point of b on $\mathcal{C}$ is either p_{uv}^b or

is located in between p_{uv}^b and a. As the order of furthest points has to be preserved in every cycle, p_{uv}^{a*} and p_{uv}^{b*} have to be located in between a and b.

Applying Remark 2.2(i) twice from "edge" $p_{uv}^a p_{uv}^b$ to ab, with points p_{uv}^{a*} and p_{uv}^{b*} in ab, yields

$$\mu_c(ab, p_{uv}^a p_{uv}^b) = \frac{|ap_{uv}^{a*}|}{|ab|} \mu_c(ap_{uv}^{a*}, p_{uv}^a p_{uv}^b) + \frac{|p_{uv}^{a*} p_{uv}^{b*}|}{|ab|} \mu_c(p_{uv}^{a*} p_{uv}^{b*}, p_{uv}^a p_{uv}^b)$$
$$+ \frac{|p_{uv}^{b*} b|}{|ab|} \mu_c(p_{uv}^{b*} b, p_{uv}^a p_{uv}^b).$$

With an analogous argument as above, $\mu_c(ap_{uv}^{a*}, p_{uv}^a p_{uv}^b)$ and $\mu_c(p_{uv}^{b*} b, p_{uv}^a p_{uv}^b)$ can be obtained by Lemma 3.3 as linear cases (for the first case, for example, take the two shortest paths connecting, respectively, p_{uv}^a and p_{uv}^b with a, which go through u and v; p_{uv}^a and p_{uv}^b satisfy the same component property with respect to ap_{uv}^{a*} and endpoint p_{uv}^{a*}).

The value $\mu_c(p_{uv}^{a*} p_{uv}^{b*}, p_{uv}^a p_{uv}^b)$ can be computed as a rectangular case of Lemma 3.2. It is easy to check that $|p_{uv}^a p_{uv}^b| = |p_{uv}^{a*} p_{uv}^{b*}|$ as the distances $d(p_{uv}^a, p_{uv}^{a*})$ and $d(p_{uv}^b, p_{uv}^{b*})$ equal the semiperimeter of $\mathcal{C}$. This also implies that the paths on $\mathcal{C}$ connecting, respectively, p_{uv}^{b*} with p_{uv}^a, and p_{uv}^{a*} with p_{uv}^b, have the same length.

In total, there are at most four linear cases and one rectangular case in order to obtain $\mu_c(ab, uv)$. If p_{uv}^a and p_{uv}^b are the same point, the number of linear cases reduces to two and there is no rectangular case since $\mu_c(ab, uv) = \frac{|up_{uv}^a|}{|uv|} \mu_c(ab, up_{uv}^a) + \frac{|p_{uv}^a v|}{|uv|} \mu_c(ab, p_{uv}^a v)$. $\qquad\square$

Next, we observe that the conditions that need to be checked to compute the continuous mean distance between two edges can be checked in constant time. Notice that we do not need the explicit construction of the continuous shortest path trees from each vertex, we only need to check the conditions in Lemmas 3.5 and 3.6 below.

Lemma 3.5. *Given $v \in V(G_\ell)$, $e = ab \in E(G_\ell)$, and the values of $d(v, b)$ and $d(v, a)$, it can be checked in constant time whether edge e belongs to $E(T_v)$.*

Proof. It follows from the fact that $e \notin E(T_v)$ if and only if $|d(v, b) - d(v, a)| < |ab|$. $\qquad\square$

Lemma 3.6. *For every edge $uv \in E(G_\ell)$, it can be checked in constant time whether u and v satisfy the same component property with respect to any other edge $ab \in E(G_\ell)$ and one of its endpoints, assuming that $d(a, u), d(a, v), d(b, u)$, and $d(b, v)$ are known.*

Proof. Consider an edge ab and the endpoint a. Having the same component property with respect to edge ab and vertex a is equivalent to say that (i) $d(a, u) = d(a, b) + d(b, u)$, and (ii) $d(a, v) = d(a, b) + d(b, v)$. $\qquad\square$

Since any pair of distinct edges falls into one of the three cases considered above (rectangular, linear, or cycle), we conclude the following.

Theorem 3.2. *Let G be a weighted graph. Given two edges $e, e' \in E(G_\ell)$, the function $\mu_c(e, e')$ can be expressed as a weighted sum of $O(1)$ distances between pairs of points on e and e'.*

3.2. A Geometric View Based on Lower Envelopes and Voronoi Diagrams

In this subsection, we present an alternative approach based on well-known geometric tools, which shows that the continuous mean distance between two edges and, therefore, of the whole graph can be computed completely using simple geometric arguments. Recall that the *lower envelope* of a set of functions is the function resulting from taking the point-wise minimum of all functions in the set.

Lemma 3.7. *Given two distinct edges $e, e' \in E(G_\ell)$, the function $d(p, q)$, where $p \in e$ and $q \in e'$ can be seen as the lower envelope of at most four planes in 3D.*

Proof. Any path connecting points p and q must go through an endpoint of $e = uv$ and an endpoint of $e' = u'v'$. Assume first that the four endpoints are distinct. Parametrizing the points on e and e' as $p = xv + (1-x)u$ and $q = yv' + (1-y)u'$ for $x, y \in [0,1]$, we obtain four planes. For each of the four possible pairs of endpoints of e and e', the corresponding plane gives the length of a shortest pq-path among the pq-paths that go through those endpoints; their equations are:

$$
\begin{aligned}
P(u, u') &: z = |e|x + |e'|y + d(u, u') \\
P(u, v') &: z = |e|x + |e'|(1 - y) + d(u, v') \\
P(v, u') &: z = |e|(1 - x) + |e'|y + d(v, u') \\
P(v, v') &: z = |e|(1 - x) + |e'|(1 - y) + d(v, v')
\end{aligned}
\tag{8}
$$

where, for instance, $P(u, u')$ indicates that the paths considered to connect the points on e with those on e' go through endpoints u and u'. Hence, for any two points $p \in e$ and $q \in e'$, the function $d(p, q)$ is the minimum among the four values obtained.

The above argument can be adapted naturally when e and e' have a common endpoint, in which case there are only two planes. $\square$

The previous result implies that the continuous mean distance between any two edges can be computed in constant time if the distances between their endpoints are known. However, we give next a direct way to compute it that avoids the computation of lower envelopes. In particular, we show that it is also possible to compute $\mu_c(e, e')$ in constant time by considering the volume of a three-dimensional body with a rectangular base (one side with the length of e and the other with the length of e'), four vertical faces from each of the four base edges, and a *roof* that is the lower envelope defined in Lemma 3.7. Next, we describe how this lower envelope or *roof* can be viewed.

We consider a rectangle whose corners are labeled with the possible combinations of endpoints of $e = uv$ and $e' = u'v'$, as done to define the four planes
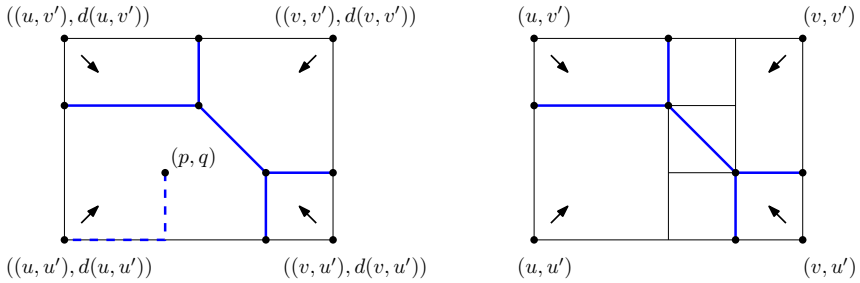
(a) The length of the dashed path allows to compute the distance $d(p, q)$.

(b) Subdivision into simpler shapes for simpler computation.

FIGURE 6. The *roof* determined by the distances between points on $e = uv$ and $e' = u'v'$, and a pair of points (p, q) located in the region associated to (u, u'). Each corner is labeled with the height of the vertical edge of that corner, and the arrows show the direction of maximum slope of each roof

in the proof of Lemma 3.7. The labels also include a weight equal to the distance between the corresponding endpoints; when no confusion may arise, we shall only indicate in the figures the weights of the corners. In addition, the rectangle is split (into at most four regions) by the orthogonal projection onto the (x, y)-plane of the (at most five) intersections of the planes defined by the equations in (8). Refer to Fig. 6a.

Thus, for instance, a pair (p, q), with $p \in e$ and $q \in e'$, is located in the region associated to (u, u') if $d(p, q)$ is given by a path that goes through u and u', that is, by the plane $z = |e|x + |e'|y + d(u, u')$ of Lemma 3.7. Hence, for $p = xv + (1 - x)u$ and $q = yv' + (1 - y)u'$ with $x, y \in [0, 1]$, the value $d(p, q)$ is just the distance in the L_1 metric[3] from (p, q) to the corner (u, u') plus the weight of that corner, which is $d(u, u')$. This is analogous for the remaining corners of the rectangle. Thus, the projections of the intersections between the planes of equations in (8) can be viewed as the bisectors of the additively weighted Voronoi diagram for the L_1 metric [24] of the corners of the rectangle. Therefore, the first step to compute $d(p, q)$ is to determine the region in which (p, q) lies, as it determines the plane that defines the lower envelope over (p, q).

The mean distance of the points in each of the (at most) four Voronoi regions is the volume of a truncated prism (with the corresponding Voronoi region as base and the corresponding plane of Lemma 3.7 as roof) divided by the area of the base. From a practical point of view, since there is no formula for a direct computation of that volume, it is better to subdivide the original Voronoi diagram into sub-rectangles and triangles, as Fig. 6b shows, since, in those cases, the volume of the truncated prism is given by the average height

---

[3]The L_1-distance between two points $p = (x_p, y_p)$ and $q = (x_q, y_q)$ is given by $|x_p - x_q| + |y_p - y_q|$.
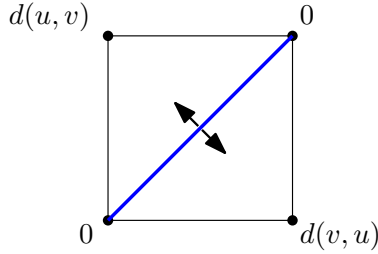
FIGURE 7. The case $e = e'$

of the corners, see [15]. In the proof of Proposition 4.3, we use this technique to compute the continuous mean distance of the 1-uniform complete graph, and it is also referred in the proof of Proposition 5.1.

Remark 3.1. Although we have already seen in Eq. (6) that $\mu_c(e, e) = |e|/3$, we can give an interpretation of this value in terms of a roof-diagram (see Fig. 7). In this case, the rectangle becomes a square, and the roof is formed by two planes:

$$P(u, v) : z = |e|(y - x)$$
$$P(v, u) : z = |e|(x - y),$$

where $P(u, v)$ indicates that point $p = xv + (1 - x)u$ is closer to u than $q = yv + (1 - y)u$ (analogous for $P(v, u)$).

Summarizing the above discussion, we have the following theorem.

Theorem 3.3. *Let G be a connected weighted graph. Given two edges $e, e' \in E(G_\ell)$, the function $\mu_c(e, e')$ can be expressed as a weighted volume of at most eight truncated rectangular prisms.*

4. Specific Cases: Trees, Cactus, and Complete Graphs

The study developed in the previous section reflects the difficulties of computing the continuous mean distance, even for specific weighted graphs. As mentioned in the Introduction, the value of this parameter is only known for seven other simple graphs and six very specific graph families [9,10]. In this section we deal with complete graphs and graphs that have cut vertices. For graphs that have this structural property, the continuous mean distance can be computed faster than using Theorem 3.1, by studying each block independently.

Lemma 4.1. *Let G_ℓ be a weighted connected graph with a cut-vertex v, i.e., $G_\ell = G_\ell^1 \cup G_\ell^2$ and $G_\ell^1 \cap G_\ell^2 = \{v\}$. Then, $\mu_c(G_\ell^1, G_\ell^2) = \mu_c(v, G_\ell^1) + \mu_c(v, G_\ell^2)$*

and

$$\mu_c(G_\ell) = \left(\frac{|G_\ell^1|}{|G_\ell|}\right)^2 \mu_c(G_\ell^1) + \left(\frac{|G_\ell^2|}{|G_\ell|}\right)^2 \mu_c(G_\ell^2)$$
$$+ 2\left(\frac{|G_\ell^1||G_\ell^2|}{|G_\ell|^2}\right)\left(\mu_c(v, G_\ell^1) + \mu_c(v, G_\ell^2)\right).$$

Proof. If $p \in G_\ell^1$ and $q \in G_\ell^2$, then $d(p,q) = d(p,v) + d(v,q)$. By Eqs. (4) and (3), we have:

$$\mu_c(G_\ell^1, G_\ell^2) \overset{(4)}{=} \frac{1}{|G_\ell^1||G_\ell^2|} \iint_{p \in G_\ell^1,\, q \in G_\ell^2} d(p,q)\, dp\, dq$$

$$= \frac{1}{|G_\ell^1||G_\ell^2|} \left(\iint_{p \in G_\ell^1,\, q \in G_\ell^2} d(p,v)\, dp\, dq + \iint_{p \in G_\ell^1,\, q \in G_\ell^2} d(v,q)\, dp\, dq\right)$$

$$\overset{(3)}{=} \frac{1}{|G_\ell^1||G_\ell^2|} \left(|G_\ell^1| \int_{q \in G_\ell^2} \mu_c(v, G_\ell^1)\, dq + |G_\ell^2| \int_{p \in G_\ell^1} \mu_c(v, G_\ell^2)\, dp\right).$$

Since in the last equation the integrated functions are constant with respect to the corresponding differentials, we obtain:

$$\mu_c(G_\ell^1, G_\ell^2) = \frac{1}{|G_\ell^1||G_\ell^2|}\left(\mu_c(v, G_\ell^1)|G_\ell^1||G_\ell^2| + \mu_c(v, G_\ell^2)|G_\ell^1||G_\ell^2|\right)$$
$$= \mu_c(v, G_\ell^1) + \mu_c(v, G_\ell^2).$$

The formula for $\mu_c(G_\ell)$ is then obtained as follows:

$$\mu_c(G_\ell) \overset{(4)}{=} \frac{1}{|G_\ell|^2} \iint_{p,q \in G_\ell} d(p,q)\, dp\, dq$$

$$= \frac{1}{|G_\ell|^2}\left(\iint_{p,q \in G_\ell^1} d(p,q)\, dp\, dq + \iint_{p,q \in G_\ell^2} d(p,q)\, dp\, dq\right.$$

$$\left. + 2\iint_{p \in G_\ell^1,\, q \in G_\ell^2} d(p,q)\, dp\, dq\right)$$

$$= \frac{1}{|G_\ell|^2}\left(|G_\ell^1|^2 \mu_c(G_\ell^1) + |G_\ell^2|^2 \mu_c(G_\ell^2) + 2|G_\ell^1||G_\ell^2|\mu_c(G_\ell^1, G_\ell^2)\right)$$

$$= \frac{1}{|G_\ell|^2}\left(|G_\ell^1|^2 \mu_c(G_\ell^1) + |G_\ell^2|^2 \mu_c(G_\ell^2) + 2|G_\ell^1||G_\ell^2|(\mu_c(v, G_\ell^1) + \mu_c(v, G_\ell^2))\right).$$

$$\square$$

Thus, if we know a direct formula to obtain the continuous mean distance of each block of G_ℓ, then $\mu_c(G_\ell)$ can be computed in linear time. For instance, this is the case for trees.

Proposition 4.1. *The continuous mean distance of a weighted tree T_ℓ with n vertices can be computed in $O(n)$ time.*

Proof. We apply induction on n. The continuous mean distance of an edge e is, by Eq. (6), $|e|/3$. For $n \geq 3$, take a non-leaf vertex v of a tree T_ℓ, which is a cut-vertex, and consider the two sub-trees connected by v. By Lemma 4.1, $\mu_c(T_\ell)$ is obtained by computing the total length of each sub-tree, its continuous mean distance, and its continuous mean distance from v. By induction, these values can be computed in linear time and combined in constant time to obtain $\mu_c(T_\ell)$. □

Another interesting application of Lemma 4.1 is for the well-known *cactus* graphs, see for instance [18,36] for studies in the context of location on graphs. This type of graphs has cut vertices, and each block is either an edge or a cycle. Since the continuous mean distance of a cycle C_ℓ is $|C_\ell|/4$ (see [10]), and that of an edge is given by Eq. (6), we obtain (again, by induction) the following result.

Proposition 4.2. *The continuous mean distance of a weighted cactus graph with n vertices can be computed in $O(n)$ time.*

When the graphs have no cut vertices, the method described in Sect. 3.2 is a useful tool to compute the continuous mean distance. Next, we apply this method to the α-uniform complete graph K_n^α. While the value $\mu_d(K_n) = (n-1)/n$ is trivial to compute, the continuous version is much harder.

Proposition 4.3. *The continuous mean distance of the α-uniform complete graph K_n^α is given by the following formula:*

$$\mu_c((K_n^\alpha)_\ell) = \frac{\alpha(9\,n^2 - 22\,n + 12)}{6\,(n^2 - n)}.$$

Proof. By Remark 2.1, it suffices to prove the result for $\alpha = 1$. We use Eq. (7) and the technique described in Sect. 3.2 to compute the continuous mean distance between two distinct edges. There are two types of pairs of distinct edges e and e', incident and non-incident:

Case 1. If e and e' are incident at a vertex u, there is another edge connecting their non-common endpoints, say v and v'. With respect to the description in Sect. 3.2, we only have two planes: $z = x + y$ (where $d(u, u') = 0$) and $z = 2 - x - y + d(v, v')$, and the corresponding roof–diagram is illustrated in Fig. 8a. The partition of the diagram into one rectangle, one square and two triangles is shown in Fig. 8b, where the number inside each region indicates the value of the volume of the truncated prism with that base; this number is given by the average height of the corners. Taking into account the corresponding areas of the base, we have $\mu_c(e, e') = \frac{1}{2} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{4}{3} + \frac{1}{4} = \frac{23}{24}$, and the total number of this type of pairs of edges is $n(n-1)(n-2)$.
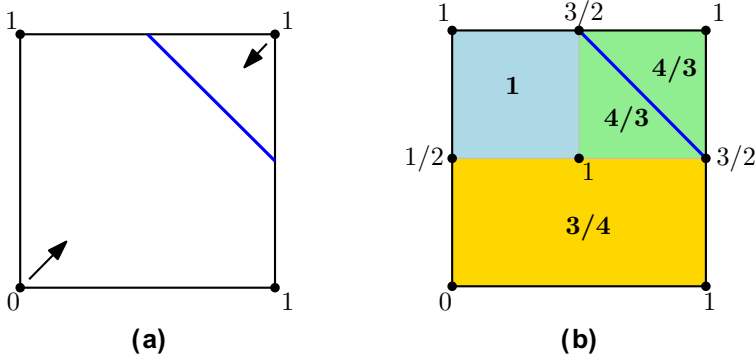
FIGURE 8. **a** Roof–diagram for two incident edges in K_n^1. **b** Partition into truncated prisms; the value of the mean distance is indicated in each region in boldface, as well as the weights of the corners
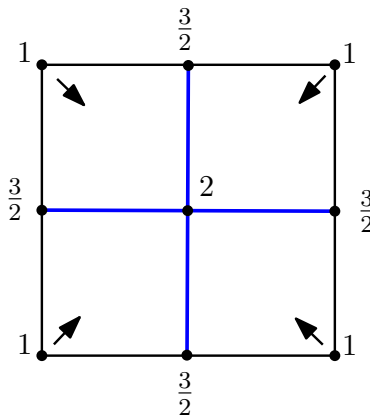

FIGURE 9. The roof–diagram for two non-incident edges in K_n^1

Case 2. If e and e' are non-incident, the roof–diagram looks as that of Fig. 9. Now, $\mu_c(e, e') = 3/2$, and the total number of these pairs of edges is

$$\binom{n}{2}^2 - n(n-1)(n-2) - \binom{n}{2},$$

where the last term is subtracted to take into account all pairs where $e = e'$, which are already considered in Eq. (7). For $\alpha = 1$, this equation then gives

$$\mu_c((K_n^1)_\ell) = \frac{4}{n^2(n-1)^2} \left(\sum_{(e,e') \in E \times E, e \neq e'} \mu_c(e, e') + \frac{n(n-1)}{6} \right)$$

where

$$\sum_{\substack{(e,e')\in E\times E \\ e\neq e}} \mu_c(e,e') = \frac{23n(n-1)(n-2)}{24}$$

$$+\frac{3}{2}\left(\frac{n^2(n-1)^2}{4} - n(n-1)(n-2) - \frac{n(n-1)}{2}\right).$$

Hence,

$$\mu_c((K_n^1)_\ell) = \frac{92(n-2)}{24n(n-1)} + \frac{3}{2} - \frac{6(n-2)}{n(n-1)} - \frac{3}{n(n-1)} + \frac{2}{3n(n-1)}$$

$$= \frac{9n^2 - 22n + 12}{6n(n-1)}.$$

□

Although this section is focused on the computation of the continuous mean distance of some specific weighted graphs, we conclude it with a result on the range of values of $\mu_c(T_\ell)$, as it extends a similar result for the discrete case, which we believe is of interest. Indeed, in [13], the author proves that the Wiener index of any tree T on n vertices is lower-bounded by the Wiener index of the star S on n vertices, and upper-bounded by the Wiener index of the path P with the same number of vertices (where the three graphs are unweighted). Therefore, by definition, $\mu_d(S) \le \mu_d(T) \le \mu_d(P)$. Next, we prove that, when the graph is uniform, these bounds also hold for the continuous case.

Proposition 4.4. *Let S_ℓ and P_ℓ be an α-uniform star and α-uniform path, respectively, on n vertices. Then,*

$$\mu_c(S_\ell) \le \mu_c(T_\ell) \le \mu_c(P_\ell)$$

for every α-uniform tree T_ℓ with n vertices.

Proof. It suffices to prove the result for $\alpha = 1$ (see Remark 2.1). We apply induction on n. Let $T'_\ell := T_\ell\backslash\{u_1\}$, $S'_\ell := S_\ell\backslash\{u_2\}$, and $P'_\ell := P_\ell\backslash\{u_3\}$ where u_i is, in each case, a leaf adjacent to a vertex v_i, $1 \le i \le 3$, of the corresponding graph. Lemma 4.1 gives expressions for $\mu_c(T_\ell)$, $\mu_c(S_\ell)$, and $\mu_c(P_\ell)$ in terms of the continuous mean distances of the corresponding edge $u_i v_i$ and, respectively, T'_ℓ, S'_ℓ, and P'_ℓ (simply set G^1_ℓ as the graph on $n-1$ vertices and G^2_ℓ as the edge $u_i v_i$). For T_ℓ we obtain:

$$\mu_c(T_\ell) = \left(\frac{n-2}{n-1}\right)^2 \mu_c(T'_\ell) + \frac{1}{3(n-1)^2} + 2\left(\frac{n-2}{(n-1)^2}\right)\left(\mu_c(v_1, T'_\ell) + \frac{1}{2}\right),$$

and analogous expressions are obtained for $\mu_c(S_\ell)$ and $\mu_c(P_\ell)$ (by simply replacing T'_ℓ and v_1 by either S'_ℓ, v_2 or P'_ℓ, v_3, respectively). Hence, by induction, it suffices to prove that $\mu_c(v_2, S'_\ell) \le \mu_c(v_1, T'_\ell) \le \mu_c(v_3, P'_\ell)$ where v_2 is the central vertex of S'_ℓ, and v_3 is an endpoint of P'_ℓ.

Given a vertex v of any 1-uniform tree T with m edges, by Eq. (3) and Lemma 3.1,

$$\mu_c(v,T) = \frac{\sum_{e \in T} \mu_c(v,e)}{m} = \frac{\frac{m}{2} + \sum_{e \in T} d(v,e)}{m},$$

since all edges in the tree belong to $E(T_v)$ and have weight 1. Therefore, $\mu_c(v,T)$ is determined by the vector $V(v,T) = (d(v,e_1), d(v,e_2), \ldots, d(v,e_m))$, where $E(T) = \{e_1, \ldots, e_m\}$ and $d(v,e_i)$ is sorted in increasing order. The first coordinate of the vector is always 0 (v belongs to at least one edge), and the difference between two consecutive coefficients of the vector is at most 1 (the length of any edge in the tree). Thus, the smallest possible vector in a tree with $n-2$ edges (ordered by the sum of its coordinates) is $(0, 0, \ldots, 0)$, which corresponds to the case of the star, and the largest possible vector is $(0, 1, 2, \ldots, n-3)$, which corresponds to the path. This implies that $\mu_c(v_2, S'_\ell) \leq \mu_c(v_1, T'_\ell) \leq \mu_c(v_3, P'_\ell)$. □

5. Discrete Versus Continuous Mean Distances

There is no obvious relation between the discrete and the continuous mean distances, in the sense that for different graphs, any of them can be larger. From the result for K_n^α in the previous section (Proposition 4.3) it follows that the continuous mean distance can be larger than the discrete counterpart. This also happens for cycles[4] and, on the other hand, we have seen (in the Introduction) that the opposite occurs for paths.

This and the following section are devoted to better understanding the relationship between the two parameters. We first present bounds on the continuous mean distance of two edges in terms of discrete distances, which lead to bounds for the whole graph (in Corollary 5.1 below) whenever it is uniform.

Proposition 5.1. *Let e and e' be two distinct edges in a weighted graph G. Then,*

$$d(e, e') + \frac{|e| + |e'|}{4} \leq \mu_c(e, e') \leq d(e, e') + \frac{|e| + |e'|}{2},$$

and both bounds are tight.

Proof. For the upper bound, let $p \in e = ab$ and $q \in e' = uv$ and, without loss of generality, let $d(e, e') = d(a, u)$. We have $d(p, q) \leq d(p, a) + d(a, u) + d(u, q) = d(p, a) + d(e, e') + d(u, q)$. Hence,

$$\mu_c(e, e') \overset{(4)}{=} \frac{1}{|e||e'|} \iint_{p \in e, \, q \in e'} d(p, q) \, dp \, dq$$

$$\leq \frac{1}{|e||e'|} \left(\iint_{p \in e, \, q \in e'} d(p, a) \, dp \, dq + \iint_{p \in e, \, q \in e'} d(e, e') \, dp \, dq \right)$$

---

[4]The continuous mean distance of an 1-uniform cycle C_n of n vertices is $n/4$ [10], as well as $\mu_d(C_n)$ for n even; otherwise $\mu_d(C_n) = \frac{n}{4} - \frac{1}{4n}$ [34].

$$+ \iint_{p \in e,\, q \in e'} d(u, q) \, dp \, dq \Bigg)$$

$$= \frac{1}{|e||e'|} \left(\int_{q \in e'} \frac{|e|^2}{2} \, dq + d(e, e')|e||e'| + \int_{p \in e} \frac{|e'|^2}{2} \, dp \right)$$

$$= \frac{|e|}{2} + d(e, e') + \frac{|e'|}{2}.$$

Note that $\int_{p \in e} d(p, a) \, dp = \frac{|e|^2}{2}$ since a is an endpoint of the edge e, and so the integral is the area of a triangle with base and height equal to $|e|$. Analogously, $\int_{q \in e'} d(u, q) \, dq = \frac{|e'|^2}{2}$.

For the lower bound we have:

$$d(p, q) = \min \left\{ \begin{array}{l} d(p, a) + d(q, u) + d(a, u) \\ d(p, b) + d(q, u) + d(b, u) \\ d(p, a) + d(q, v) + d(a, v) \\ d(p, b) + d(q, v) + d(b, v) \end{array} \right\} \geq \min \left\{ \begin{array}{l} d(p, a) + d(q, u) \\ d(p, b) + d(q, u) \\ d(p, a) + d(q, v) \\ d(p, b) + d(q, v) \end{array} \right\} + d(e, e').$$

If the last minimum is denoted by $\Theta(p, q)$, then $d(p, q) \geq \Theta(p, q) + d(e, e')$. Therefore,

$$\mu_c(e, e') \overset{(4)}{=} \frac{1}{|e||e'|} \iint_{p \in e,\, q \in e'} d(p, q) \, dp \, dq$$

$$\geq \frac{1}{|e||e'|} \left(\iint_{p \in e,\, q \in e'} \Theta(p, q) \, dp \, dq + \iint_{p \in e,\, q \in e'} d(e, e') \, dp \, dq \right)$$

$$= \frac{1}{|e||e'|} \left(|e||e'| \frac{|e| + |e'|}{4} + |e||e'|d(e, e') \right) = \frac{|e| + |e'|}{4} + d(e, e'),$$

where $\iint_{p \in e,\, q \in e'} \Theta(p, q) \, dp \, dq$ is the volume determined by the roof–diagram depicted in Fig. 10, which can be computed as follows:

$$\iint_{p \in e,\, q \in e'} \Theta(p, q) \, dp \, dq = 4 \iint_{p \in [a, \frac{a+b}{2}],\, q \in [u, \frac{u+v}{2}]} d(p, a) + d(q, u) \, dp \, dq$$

$$= 4 \left(\frac{|e|^2}{8} \cdot \frac{|e'|}{2} + \frac{|e'|^2}{8} \cdot \frac{|e|}{2} \right) = |e||e'| \left(\frac{|e| + |e'|}{4} \right).$$

Next, we observe that both bounds are tight. For instance, the mean distance of two edges that are connected by a unique path gives the upper bound, and the lower bound is attained by two edges whose endpoints are at the same distance so that $d(e, e')$ is given by any of the four possible combinations of endpoints. □

As a consequence of the preceding proposition we obtain, for α-uniform graphs G, bounds on $\mu_c(G_\ell)$ in terms of the discrete mean distance of a weighted version of its line graph. Recall that the *line graph* $L(G)$ of an unweighted graph G has a vertex associated with each edge in G, and two vertices
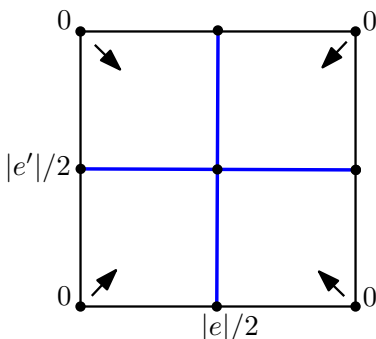
FIGURE 10. The roof–diagram for $\Theta(p, q)$

are adjacent if the corresponding edges of G have a vertex in common. When G is α-uniform, we consider the α-*uniform* line graph $L_\alpha(G)$ that is defined analogously but, in addition, every edge has length α.

The Wiener index of $L(G)$ is known as the *edge-Wiener index* of G (see, for instance, [7,28] and the references therein). It is defined as $W_e(G) = \sum_{\{e,e'\}\subseteq E} d_{L(G)}(e, e')$, where $d_{L(G)}(e, e')$ is the distance of the corresponding vertices (to e and e') in $L(G)$. This can be naturally extended to $L_\alpha(G)$ for α-uniform graphs G with m edges, and thus we may consider the discrete mean distance $\mu_d(L_\alpha(G)) = 2W(L_\alpha(G))/m^2$.

Corollary 5.1. *Let G be an α-uniform graph with m edges. Then,*

$$\mu_d(L_\alpha(G)) + \frac{\alpha}{3m} - \frac{(m-1)\alpha}{2m} \leq \mu_c(G_\ell) \leq \mu_d(L_\alpha(G)) + \frac{\alpha}{3m}$$

where $L_\alpha(G)$ is the α-uniform line graph of G.

Proof. It can be easily checked that, by construction, $d(e, e') = d_{L(G)}(e, e') - 1$ for distinct edges e and e' in an unweighted graph G; this extends to $d(e, e') = d_{L_\alpha(G)}(e, e') - \alpha$ when the graph G is α uniform. Proposition 5.1 then gives $d_{L_\alpha(G)}(e, e') - \frac{\alpha}{2} \leq \mu_c(e, e') \leq d_{L_\alpha(G)}(e, e')$ for distinct edges e and e' of G_ℓ. Now, by Eq. (7), we obtain:

$$\mu_c(G_\ell) = \frac{1}{\alpha^2 m^2} \left(\sum_{(e,e')\in E\times E, e\neq e'} \mu_c(e, e')\alpha^2 + \sum_{e\in E} \frac{\alpha^3}{3} \right)$$

$$= \frac{1}{m^2} \sum_{(e,e')\in E\times E, e\neq e'} \mu_c(e, e') + \frac{\alpha}{3m}.$$

Therefore,

$$\mu_c(G_\ell) \geq \frac{1}{m^2} \left(\sum_{(e,e')\in E\times E, e\neq e'} d_{L_\alpha(G)}(e, e') \right) + \frac{\alpha}{3m} - \frac{(m-1)\alpha}{2m}$$

and

$$\mu_c(G_\ell) \leq \frac{1}{m^2}\left(\sum_{(e,e')\in E\times E, e\neq e'} d_{L_\alpha(G)}(e,e')\right) + \frac{\alpha}{3m}.$$

The result then follows by definition of $\mu_d(L_\alpha(G))$. $\square$

In their seminal work on the mean distance for shapes [9], Doyle and Graver already studied the continuous mean distance by using an iterative edge refinement process. Following this direction, next we present an edge subdivision approach for trees. One may think that if the subdivision points are chosen arbitrarily, the discrete mean distance of the refined tree may increase significantly with respect to the original one. However, this is not the case, as we can see in the next theorem.

Theorem 5.1. *Let T be a weighted tree with n vertices, and let $T^{(k)}$ be the tree resulting from subdividing each edge of T by adding k new vertices on it. Then*
$$\mu_d(T^{(k)}) < \frac{n}{n - \frac{2k}{k+1}}\mu_d(T).$$

Proof. We start by noting that $T^{(k)}$ has $n + k(n-1)$ vertices: n *old* vertices and $k(n-1)$ *new* vertices. The value $\mu_d(T^{(k)})$ is the average of $(n+k(n-1))^2$ distances, of three types: (i) between two old vertices, (ii) between an old and a new vertex, (iii) between two new vertices. To avoid any confusion, we shall use d_T and $d_{T^{(k)}}$ to indicate distances in, respectively, the trees T and $T^{(k)}$. Further, with some abuse of notation, we shall write $u \in T$ instead of $u \in V(T)$.

Consider a distance of type (ii), between an old vertex $u \in T$ and a new vertex $a \in T^{(k)}\setminus T$. Since $T^{(k)}$ is a tree, there is a unique path from u to a. Moreover, since a is interior to an edge of T, the path can be extended in direction away from u until the first old vertex $v \in T$. Clearly, $d_{T^{(k)}}(u,a) \leq d_{T^{(k)}}(u,v) = d_T(u,v)$.

Similarly, associated to a distance of type (iii), between two new vertices $a, b \in T^{(k)}\setminus T$, there is a unique path in $T^{(k)}$ that can be extended in both directions until starting and ending, respectively, at vertices $u, v \in T$. Again, $d_{T^{(k)}}(a,b) \leq d_{T^{(k)}}(u,v) = d_T(u,v)$.

In this way, each distance involving a new vertex (types ii or iii) can be upper-bounded by a distance between two old vertices. Moreover, the distance $d(u,v)$ of a pair of vertices $(u,v) \in T^2$ can only be an upper bound for up to $(k^2 + 2k)$ distances: $2k$ of type (ii), and k^2 of type (iii). Observe that the same happens for the distance $d(v,u)$. This leads to an upper-bound on $\mu_d(T^{(k)})$ as follows:

$$\mu_d(T^{(k)})$$
$$= \frac{\sum\limits_{(u,v)\in T^2} d_{T^{(k)}}(u,v) + \sum\limits_{(u,a)\in T\times(T^{(k)}\setminus T)} d_{T^{(k)}}(u,a) + \sum\limits_{(a,u)\in(T^{(k)}\setminus T)\times T} d_{T^{(k)}}(a,u) + \sum\limits_{(a,b)\in(T^{(k)}\setminus T)^2} d_{T^{(k)}}(a,b)}{(n(k+1)-k)^2}$$

$$\leq \frac{\sum\limits_{(u,v)\in T^2} d_T(u,v) + (k^2+2k)\sum\limits_{(u,v)\in T} d_T(u,v)}{(n(k+1)-k)^2} = \frac{(k+1)^2 \sum\limits_{(u,v)\in T^2} d_T(u,v)}{(n(k+1)-k)^2}.$$

Since $(n(k+1)-k)^2 = n^2(k+1)^2(1-\frac{2k}{n(k+1)})+k^2 > n^2(k+1)^2(1-\frac{2k}{n(k+1)})$, we obtain:

$$\mu_d(T^{(k)}) < \frac{(k+1)^2 \sum\limits_{(u,v)\in T^2} d_T(u,v)}{n^2(k+1)^2(1-\frac{2k}{n(k+1)})} = \frac{\mu_d(T)}{1-\frac{2k}{n(k+1)}} = \frac{n}{n-\frac{2k}{k+1}}\mu_d(T).$$

$\square$

Theorem 5.1 gives an upper bound of $\frac{n}{n-2}\mu_d(T)$ for the discrete mean distance of any subdivision of a tree T but, if the subdivision points are chosen more carefully, one can expect more precise results. Thus, Theorem 5.1 is the initial motivation of the following section where, in particular, we explore the convergence of the discrete mean distance to its continuous counterpart when subdividing the edges of a tree (see Corollary 6.2).

6. Convergence: Graph Subdivision

A natural question is whether the discrete mean distance is convergent to its continuous counterpart when iteratively subdividing the edges of the graph. One may propose different subdivision schemes but, as we shall see later in this section, not all of them guarantee convergence. By definition of continuous mean distance, the convergence happens for uniform graphs by simply adding, at each step, a new vertex on each edge. Further, it is not hard to devise a subdivision scheme with guaranteed convergence if the ratio between the lengths of the longest and the shortest edges approaches one as the subdivision progresses. However, such a scheme completely depends on the original structure of the graph.

In this section we present an edge subdivision scheme that does not depend on the graph structure, and allows us to obtain bounds on the discrete mean distance of its k-th edge subdivision, and on its limit when k tends to infinity. We begin by introducing some notation.

For a graph $G = (V, E)$ with n vertices and m edges, let $G^1 = (V^1, E^1)$ be the graph that results from subdividing each edge of G by inserting a new vertex at its midpoint. Then, for a given $k \geq 2$, we subdivide each edge of G^1 into 2^{k-1} new edges of the same length by inserting $2^{k-1} - 1$ vertices. The resulting graph $G^k = (V^k, E^k)$ is called the *k-th subdivision of* G (note that the graph G^1 could be viewed as a subdivision of G but for our purpose it will be distinguished). Refer to Fig. 11. The vertices of the original graph G are called *black vertices*; the set of vertices inserted into G to obtain G^1 is denoted by $\mathcal{B}$, and they are called *blue vertices*. Thus, $V^1 = V \cup \mathcal{B}$ and $|\mathcal{B}| = m$. We use $\mathcal{R}^k$ to refer to the set of new vertices inserted into G^1, which are called the *red*
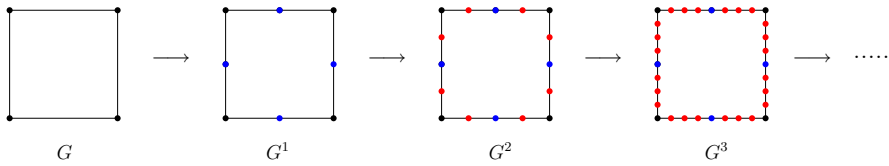
FIGURE 11. Subdividing the edges of a graph

vertices; clearly, $|\mathcal{R}^k| = 2m(2^{k-1} - 1) = m(2^k - 2)$. Hence, $V^k = V \cup \mathcal{B} \cup \mathcal{R}^k$ and $|V^k| = n + m(2^k - 1)$. Further, the edges of the original graph G can be identified within G^k: we write e^k to indicate the k-th subdivision of an edge $e \in E$, which is a path in G^k with $2^k + 1$ vertices (there are 2 black, 1 blue, and $2^k - 2$ red vertices). Equation (2) then gives:

$$\mu_d(G^k) = \frac{2W(G^k)}{(n + m(2^k - 1))^2}, \tag{9}$$

where $W(G^k) = \sum_{\{u,v\} \subset V^k} d(u,v)$ is the Wiener index of G^k. With some abuse of notation, we write, for sets $A, B \subseteq V^k$, $W(A; B) = \sum_{u \in A \backslash B,\, v \in B \backslash A} d(u,v) + \sum_{\{u,v\} \subseteq A \cap B} d(u,v)$ and $W(A) = W(A; A)$. Further, we shall indistinctly use sets of vertices or graphs in this notation, for instance, $W(A; e^k)$ is simply a sum of distances between vertices in the set A and vertices in the path e^k. With this notation, for $k \geq 2$, we have:

$$W(G^k) = W(V^1) + W(\mathcal{R}^k) + W(\mathcal{R}^k; V) + W(\mathcal{R}^k; \mathcal{B}) \tag{10}$$

    The following two subsections are devoted to proving our main result in this section (which, in addition, will allow us to gain a deeper understanding of the limit of $\mu_d(G^k)$ when k tends to infinity):

Theorem 6.1. *Let $G = (V, E)$ be a weigthed graph with n vertices and $m \geq 2$ edges, and let $G^k = (V^k, E^k)$ be its k-th subdivision, where $k \geq 2$. Let $\mathcal{B}$ be the set of vertices inserted into G to obtain the graph G^1. Then,*

$$\frac{2\left[\Omega_k(G, G^1) - \rho\left(3\binom{m}{2} + m(n-2)\right)\left(2^{k-2} - \frac{1}{2}\right)\right]}{(n + m(2^k - 1))^2} < \mu_d(G^k) \leq \frac{2\,\Omega_k(G, G^1)}{(n + m(2^k - 1))^2}$$

where $\rho = \max\{|e| : e \in E\}$, and

$$\Omega_k(G, G^1) = W(V^1) + (2^k - 2)\left(2^k W(\mathcal{B}) + W(\mathcal{B}; V)\right) + |E|\left(\frac{2^{2k-1}}{3} - 2^{k-1} + \frac{1}{3}\right).$$

Moreover, the upper bound is tight.

    The limit of the upper bound in Theorem 6.1, when k tends to infinity, is given by the coefficients of the term 2^{2k}. Thus,

$$\lim_{k \to \infty} \mu_d(G^k) \leq \lim_{k \to \infty} \frac{2\,\Omega_k(G, G^1)}{(n + m(2^k - 1))^2} = \frac{2(W(\mathcal{B}) + |E|/6)}{m^2}.$$

Hence, by Eq. (2), we obtain the following bounds.

Corollary 6.1. *Let $G = (V, E)$ be a weighted graph with $m \geq 2$ edges, and let $G^k = (V^k, E^k)$ be its k-th subdivision, where $k \geq 2$. Let $\mathcal{B}$ be the set of vertices inserted into G to obtain the graph G^1. Then,*

$$\lim_{k \to \infty} \mu_d(G^k) \leq \mu_d(\mathcal{B}) + \frac{|E|}{3m^2}.$$

Moreover, if G is α-uniform then,

$$\lim_{k \to \infty} \mu_d(G^k) = \mu_c(G_\ell) \leq \mu_d(\mathcal{B}) + \frac{\alpha}{3m}.$$

We want to highlight that all trees attain the preceding upper bounds (this is a consequence of the study developed in Sect. 6.1 below).

Corollary 6.2. *Let $T = (V, E)$ be a weighted tree with $n \geq 3$ vertices, and let T^k be its k-th subdivision. Let $\mathcal{B}$ be the set of vertices inserted into T to obtain T^1. Then,*

$$\lim_{k \to \infty} \mu_d(T^k) = \mu_d(\mathcal{B}) + \frac{|E|}{3(n-1)^2}.$$

Moreover, if T is α-uniform then,

$$\mu_c(T_\ell) = \lim_{k \to \infty} \mu_d(T^k) = \mu_d(\mathcal{B}) + \frac{\alpha}{3(n-1)}.$$

Corollary 6.2 gives simple examples where the discrete mean distance, when subdividing the edges of the graph, does not converge to the continuous counterpart. Consider, for example, a path P with 4 vertices and edge lengths $2, 1, 1$; we have $\mu_c(P_\ell) = 4/3 \approx 1.33$, whilst from Corollary 6.2 we obtain $\lim_{k \to \infty} \mu_d(P^k) = 10/9 + 4/27 = 34/27 \approx 1.26$.

6.1. Proof of the Upper Bound in Theorem 6.1

First, we upper bound the Wiener index of G^k. By Equation (10), it suffices to compute upper bounds on $W(\mathcal{R}^k), W(\mathcal{R}^k; V)$, and $W(\mathcal{R}^k; \mathcal{B})$.

To upper-bound $W(\mathcal{R}^k)$, we begin by distinguishing the distances between red vertices, depending on whether they are on the k-th subdivision of the same edge or of distinct edges. We use the notation $e^k \cap \mathcal{R}^k$ to indicate the set of red points that are on the k-th subdivision of an edge e; analogously, the notation $e^k \cap \mathcal{B}$ will refer to the blue point that is on e^k.

$$W(\mathcal{R}^k) = \sum_{\substack{\{e_1, e_2\} \subseteq E \\ e_1 \neq e_2}} W(e_1^k \cap \mathcal{R}^k; e_2^k \cap \mathcal{R}^k) + \sum_{e \in E} W(e^k \cap \mathcal{R}^k). \qquad (11)$$

For red vertices that are on the same edge, we can compute the sum of distances exactly.

Lemma 6.1. *The following formula holds for $k \geq 2$:*

$$\sum_{e \in E} W(e^k \cap \mathcal{R}^k) = |E| \left(\frac{2^{2k-1}}{3} - 3 \cdot 2^{k-2} + \frac{5}{6} \right).$$

Proof. Given an edge $e = wz \in E$, we have:

$$W(e^k \cap \mathcal{R}^k) = W(e^k) - W(\{w, z\}; e^k) - W(e^k \cap \mathcal{R}^k; e^k \cap \mathcal{B}). \qquad (12)$$

The value $W(e^k)$ is the Wiener index of an $\frac{|e|}{2^k}$-uniform path on $2^k + 1$ vertices, which can be deduced from $\mu_d(e^k)$ by Eq. (2) and Remark 2.1. It is also known that the discrete mean distance of a 1-uniform path on n vertices is $(n+1)(n-1)/3n$ [34]. Thus, we have:

$$W(e^k) = \frac{|e|}{2^k} \left(\frac{\mu_d(e^k)(2^k + 1)^2}{2} \right) = \frac{|e|}{2^k} \left(\frac{(2^k + 2)(2^k + 1)2^k}{6} \right) = \frac{|e|}{2^k} \binom{2^k + 2}{3}.$$

$$(13)$$

Further,

$$W(\{w, z\}; e^k) = d(w, z) + \sum_{\substack{u \in \{w,z\} \\ v \in e^k - \{w,z\}}} d(u, v)$$

$$= |e| + 2 \left(\frac{|e|}{2^k}(1 + 2 + 3 + \cdots + 2^k - 1) \right) = 2^k |e|. \qquad (14)$$

Finally,

$$W(e^k \cap \mathcal{R}^k; e^k \cap \mathcal{B}) = 2 \left(\frac{|e|}{2^k}(1 + 2 + 3 + \cdots + (2^{k-1} - 1)) \right) = |e| \left(2^{k-2} - \frac{1}{2} \right).$$

$$(15)$$

By Eqs. (12)–(15) we obtain:

$$\sum_{e \in E} W(e^k \cap \mathcal{R}^k) = |E| \left(\frac{1}{2^k} \binom{2^k + 2}{3} - 2^k - 2^{k-2} + \frac{1}{2} \right)$$

$$= |E| \left(\frac{2^{2k-1}}{3} - 3 \cdot 2^{k-2} + \frac{5}{6} \right).$$

$$\square$$

It remains to upper-bound the sum of distances, for every pair of distinct edges, of red vertices that are on the subdivision of the edges; see Eq. (11). Roughly speaking, the maximum value of this sum is obtained when, for every pair of distinct edges, all shortest path between any two of their vertices use the same endpoints of those edges; see Fig. 12a, which considers the subdivisions e_1^3 and e_2^3 of two distinct edges e_1 and e_2. This is because when we can enter and get out of the edges using different endpoints, the distance between any two vertices (black, red, or blue) on the subdivision of the edges, in the best case, decreases. For example, in Fig. 12a, $d(r_6, s_5)$ would be smaller if there

would be another shortest path (not only the orange one) connecting the two edges via the endpoints z_1 and z_2. The following lemma gives the maximum value of that sum of distances.

Lemma 6.2. *If for every two distinct edges e_1 and e_2 all shortest paths connecting any two vertices located on e_1^k and e_2^k, respectively, go through the same endpoints of e_1 and e_2, then:*

$$\sum_{\{e_1,e_2\}\in E, e_1\neq e_2} W(e_1^k\cap\mathcal{R}^k, e_2^k\cap\mathcal{R}^k) = (2^k-2)^2 W(\mathcal{B}).$$

Proof. Consider two distinct edges $e_1 = w_1z_1$ and $e_2 = w_2z_2$, and assume that all shortest paths connecting any two vertices on e_1^k and e_2^k go through w_1 and w_2 (the argument is analogous for the other combinations of endpoints); it might happen that $w_1 = w_2$. Let $\{w_1, r_1, \ldots r_{2^{k-1}-1}, b_1, r_{2^{k-1}}, \ldots r_{2^k-2}, z_1\}$ be the sequence of vertices in e_1^k ordered from the leftmost to the rightmost vertex, where $b_1 \in \mathcal{B}$ and $r_i \in \mathcal{R}^k$, and let $\{w_2, s_1, \ldots s_{2^{k-1}-1}, b_2, s_{2^{k-1}}, \ldots s_{2^k-2}, z_2\}$ be the analogous sequence of vertices in e_2^k. Refer to Fig. 12a. We have:

$$d(r_i, s_j) = \begin{cases} d(b_1,b_2) - (2^{k-1}-i)\dfrac{|e_1|}{2^k} & \\ \quad -(2^{k-1}-j)\dfrac{|e_2|}{2^k} & \text{if } 1 \leq i,j \leq 2^{k-1}-1 \\ d(b_1,b_2) + (i-2^{k-1}+1)\dfrac{|e_1|}{2^k} & \\ \quad +(j-2^{k-1}+1)\dfrac{|e_2|}{2^k} & \text{if } 2^{k-1} \leq i,j \leq 2^k-2 \\ d(b_1,b_2) - (2^{k-1}-i)\dfrac{|e_1|}{2^k} & \\ \quad +(j-2^{k-1}+1)\dfrac{|e_2|}{2^k} & \text{if } 1 \leq i \leq 2^{k-1}-1, 2^{k-1} \leq j \leq 2^k-2 \\ d(b_1,b_2) + (i-2^{k-1}+1)\dfrac{|e_1|}{2^k} & \\ \quad -(2^{k-1}-j)\dfrac{|e_2|}{2^k} & \text{if } 1 \leq j \leq 2^{k-1}-1, 2^{k-1} \leq i \leq 2^k-2 \end{cases}$$

Hence, $W(e_1^k\cap\mathcal{R}^k, e_2^k\cap\mathcal{R}^k) = \sum_{r_i\in e_1^k\cap\mathcal{R}^k, s_j\in e_2^k\cap\mathcal{R}^k} d(r_i,s_j) = (2^k-2)^2 d(b_1,b_2)$, since all the expressions depending on $\frac{|e_1|}{2^k}$ and $\frac{|e_2|}{2^k}$ cancel each other out (they cancel out in pairs, for example, the expressions in $d(r_1,s_1)$ cancel out with the ones in $d(r_{2^k-2}, s_{2^k-2})$). The result then follows by summing over all pairs of distinct edges e_1 and e_2. $\qquad\square$

As explained before, by Eq. (11), and Lemmas 6.1 and 6.2, we obtain an upper bound on $W(\mathcal{R}^k)$ for the k-th subdivision of any weighted graph (with at least two edges):

$$W(\mathcal{R}^k) \leq (2^k-2)^2 W(\mathcal{B}) + |E|\left(\frac{2^{2k-1}}{3} - 3\cdot 2^{k-2} + \frac{5}{6}\right). \qquad (16)$$

An upper bound on $W(\mathcal{R}^k; V)$ is obtained by using similar arguments as in Lemma 6.2. Here we assume that for every edge $e = wz$ of G, all shortest paths connecting any vertex on e^k with any vertex in $V\backslash\{w,z\}$ go through
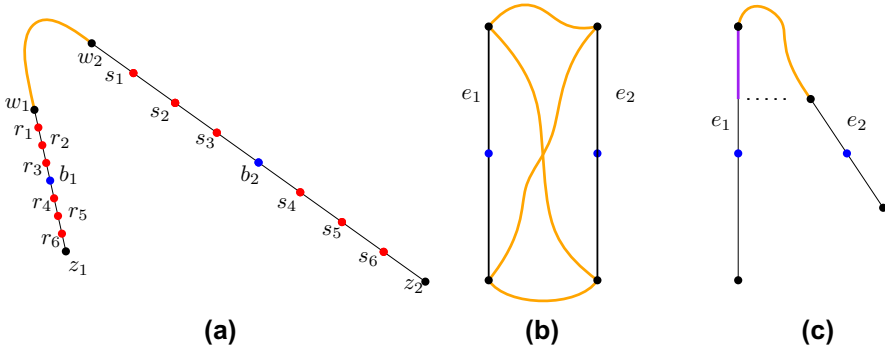
FIGURE 12. **a** Subdivisions e_1^3 and e_2^3: the shortest path between any two vertices (black, red, or blue) located on, respectively, each subdivision, goes through the orange path. **b** The orange paths are shortest paths connecting the blue points (midpoints of e_1 and e_2). **c** The shortest path between any point on the sub-edge of e_1 in purple uses the orange path, i.e., the same endpoints of e_1 and e_2; this is the condition used in Lemma 6.2 that increases the value of the sum of distances between vertices, respectively, on e_1^k and e_2^k (color figure online)

the same endpoint w of e. The only difference with the proof of Lemma 6.2 is that we compute $d(r_i, v)$ for $v \in V \setminus \{w, z\}$ instead of $d(r_i, s_j)$, obtaining analogous expressions but distinguishing only the cases $1 \leq i \leq 2^{k-1} - 1$ and $2^{k-1} \leq i \leq 2^k - 2$. The same type of expression is obtained for the endpoints of the edge e. For example, $d(r_i, w)$ is given by:

$$d(r_i, w) = \begin{cases} d(b, w) - (2^{k-1} - i)\dfrac{|e|}{2^k} & \text{if } 1 \leq i \leq 2^{k-1} - 1 \\ d(b, w) + (i - 2^{k-1} + 1)\dfrac{|e|}{2^k} & \text{if } 2^{k-1} \leq i \leq 2^k - 2 \end{cases}$$

where $b \in e^k \cap \mathcal{B}$. Thus, for a fixed $v \in V$ it follows that $\sum_{r_i \in e^k \cap \mathcal{R}^k} d(r_i, v) = (2^k - 2)d(b, v)$ (all the expressions depending on $\frac{|e|}{2^k}$ again cancel each other out). Therefore,

$$W(\mathcal{R}^k; V) = \sum_{r_i \in \mathcal{R}^k, v \in V} d(r_i, v) \leq (2^k - 2)W(\mathcal{B}; V). \tag{17}$$

and the bound is attained when the condition on the shortest paths stated above holds.

By distinguishing again between vertices that are on the same edge or on distinct edges, and proceeding as above, we reach the following upper bound on $W(\mathcal{R}^k; \mathcal{B})$. For the sake of brevity, we omit the details as the arguments

are the same.

$$W(\mathcal{R}^k; \mathcal{B}) \leq 2(2^k - 2)W(\mathcal{B}) + |E|\left(2^{k-2} - \frac{1}{2}\right). \tag{18}$$

The preceding bound is attained when, for every edge e of G, all shortest paths connecting any vertex on e^k with any vertex in $\mathcal{B} \setminus \{b\}$ (where $b \in e^k$) go through the same endpoint of e.

By Eq. (10), and the bounds given in (16), (17), and (18), we can conclude that $W(G^k) \leq \Omega_k(G, G^1)$; the upper bound in Theorem 6.1 then follows by Equation (9). This bound is attained by all weighted graphs satisfying the conditions on shortest paths that lead to the equality in Eqs. (16–18), in particular all trees.

6.2. Proof of the Lower Bound in Theorem 6.1

The minimum value of $\mu_d(G^k)$ would be obtained by a graph G satisfying that every pair of edges e_1, e_2 have the same length, and their midpoints are connected by shortest paths going through any pair of endpoints of e_1 and e_2 (see Fig. 12b). Indeed, as we explained for the upper bound, the sum of distances between vertices on the subdivisions of the edges (black, red, or blue) decreases when the combinations of endpoints to enter and get out of the edges increase, so the minimum is given when all possible combinations of endpoints can be used. In addition, the graph should be uniform, as otherwise there would be a pair of edges in the situation described in Fig. 12c, which would give a larger value for the sum of distances.

Clearly, there cannot exist a graph satisfying the previous condition on the shortest paths connecting the midpoints of any pair of edges (simply consider the midpoints of two incident edges), but, in order to make the computations necessary to obtain a lower bound on $\mu_d(G^k)$, we shall assume in Lemma 6.3 below that all edges of the graph G have the same length α, and that any pair of its edges satisfies the condition on the midpoints.

By Eq. (10), it suffices to compute lower bounds on $W(\mathcal{R}^k)$, $W(\mathcal{R}^k; V)$, and $W(\mathcal{R}^k; \mathcal{B})$ in order to lower-bound the Wiener index of G^k. Again, we begin with $W(\mathcal{R}^k)$.

Lemma 6.3. *Let G be an α-uniform graph with $m \geq 2$ edges. If the midpoints of every pair e_1, e_2 of distinct edges of G are connected by shortest paths going through any pair of endpoints of e_1 and e_2, then:*

$$\sum_{\{e_1, e_2\} \in E, e_1 \neq e_2} W(e_1^k \cap \mathcal{R}^k, e_2^k \cap \mathcal{R}^k) = (2^k - 2)^2 W(\mathcal{B}) - 2\alpha\binom{m}{2}\left(2^{k-2} - \frac{1}{2}\right).$$

Proof. Let $e_1 = w_1 z_1$ and $e_2 = w_2 z_2$, and let $b_1, b_2 \in \mathcal{B}$ be their corresponding midpoints. We follow the same notation as in the proof of Lemma 6.2 where $\{w_1, r_1, \ldots r_{2^{k-1}-1}, b_1, r_{2^{k-1}}, \ldots r_{2^k-2}, z_1\}$ and $\{w_2, s_1, \ldots s_{2^{k-1}-1}, b_2, s_{2^{k-1}}, \ldots s_{2^k-2}, z_2\}$ are the ordered sequence of vertices in, respectively, e_1^k and e_2^k. Thus,

$$d(r_i, s_j) = \begin{cases} d(b_1, b_2) - (2^{k-1} - i)\dfrac{\alpha}{2^k} & \\ \quad -(2^{k-1} - j)\frac{\alpha}{2^k} & \text{if } 1 \leq i, j \leq 2^{k-1} - 1 \\ d(b_1, b_2) - \left(i - 2^{k-1} + 1\right)\dfrac{\alpha}{2^k} & \\ \quad -(j - 2^{k-1} + 1)\frac{\alpha}{2^k} & \text{if } 2^{k-1} \leq i, j \leq 2^k - 2 \\ d(b_1, b_2) - (2^{k-1} - i)\dfrac{\alpha}{2^k} & \\ \quad -(j - 2^{k-1} + 1)\frac{\alpha}{2^k} & \text{if } 1 \leq i \leq 2^{k-1} - 1, 2^{k-1} \leq j \leq 2^k - 2 \\ d(b_1, b_2) - \left(i - 2^{k-1} + 1\right)\dfrac{\alpha}{2^k} & \\ \quad -(2^{k-1} - j)\frac{\alpha}{2^k} & \text{if } 1 \leq j \leq 2^{k-1} - 1, 2^{k-1} \leq i \leq 2^k - 2 \end{cases}$$

Hence, we obtain:

$$W(e_1^k \cap \mathcal{R}^k, e_2^k \cap \mathcal{R}^k) = (2^k - 2)^2 d(b_1, b_2) \\ - \left(W(e_1^k \cap \mathcal{R}^k, e_1^k \cap \mathcal{B}) + W(e_2^k \cap \mathcal{R}^k, e_2^k \cap \mathcal{B})\right),$$

which by Eq. (15) equals $(2^k - 2)^2 d(b_1, b_2) - 2\alpha(2^{k-2} - 1/2)$. When considering all pairs of distinct edges, the desired formula is obtained. $\square$

As it was explained before, there is no graph satisfying the conditions of Lemma 6.3, but it yields, together with Eq. (11) and Lemma 6.1, a lower bound on $W(\mathcal{R}^k)$ by considering $\rho = \max\{|e| : e \in E\}$.

$$W(\mathcal{R}^k) \geq (2^k - 2)^2 W(\mathcal{B}) - 2\rho \binom{m}{2}\left(2^{k-2} - \frac{1}{2}\right)$$
$$+ |E|\left(\frac{2^{2k-1}}{3} - 3 \cdot 2^{k-2} + \frac{5}{6}\right). \tag{19}$$

We apply similar arguments to bound $W(\mathcal{R}^k; V)$ and $W(\mathcal{R}^k; \mathcal{B})$. In both cases we distinguish whether the vertices are on the same edge or on distinct edges. For all pairs of distinct edges e_1 and e_2 (all edges of the same length α), we also assume that their midpoints are connected by shortest paths going through any pair of endpoints of e_1 and e_2. For $W(\mathcal{R}^k; V)$ we have:

$$\sum_{r_i \in e_1^k \cap \mathcal{R}^k} d(r_i, v) = (2^k - 2)d(b_1, v) - 2\frac{\alpha}{2^k}(1 + 2 + 3 + \cdots + (2^{k-1} - 1))$$

$$= (2^k - 2)d(b, v) - \alpha(2^{k-2} - 1/2)$$

where $v \in e_2^k \cap V$ and $b_1 \in e_1^k \cap \mathcal{B}$. Further, $W(e_1^k \cap \mathcal{R}^k; e_1^k \cap V) = \alpha(2^k - 2)$. By considering again $\rho = \max\{|e| : e \in E\}$ as done for Eq. (19), we obtain:

$$W(\mathcal{R}^k; V) \geq (2^k - 2)(W(\mathcal{B}; V) - m\rho) - \rho m(n - 2)\left(2^{k-2} - \frac{1}{2}\right) + m\rho(2^k - 2). \tag{20}$$

An analogous process gives $W(e_1^k \cap \mathcal{R}^k, e_2^k \cap \mathcal{B}) = (2^k - 2)d(b_1, b_2) - \alpha(2^{k-2} - 1/2)$, and together with Eq. (15) leads to:

$$W(\mathcal{R}^k; \mathcal{B}) \geq 2(2^k - 2)W(\mathcal{B}) + \left[|E| - \rho\binom{m}{2}\right]\left(2^{k-2} - \frac{1}{2}\right). \tag{21}$$

Equation (10), and the bounds in (19), (20), and (21) imply:

$$W(G^k) \geq \Omega_k(G, G^1) - \rho \left(3 \binom{m}{2} + m(n-2) \right) \left(2^{k-2} - \frac{1}{2} \right).$$

The lower bound in Theorem 6.1 again follows from Eq. (9).

7. Conclusions and Future Work

In this work we have presented the first thorough study of the continuous mean distance, a natural graph parameter that has received little attention until now. From a computational perspective, we presented two different methods to compute the mean distance of a weighted graph in roughly quadratic time in the number of edges. In addition, we obtained several structural results that provide a deeper understanding of this parameter, and can also be used to compute the mean distance faster for several graph classes. Finally, we studied the relation between the discrete mean distance and the continuous counterpart, in order to understand how the iterative subdivision of edges makes the discrete mean distance converge to the continuous one.

We are left with many intriguing questions for future research. The computational complexity of the continuous mean distance is far from settled. An important question is for what other graph classes the continuous mean distance can be computed in subquadratic time. In the case of the discrete mean distance, this was recently shown to be possible for planar graphs [4], so it is worth studying if similar techniques could be applied to the continuous setting. If that is not possible, one can still resort to approximation algorithms. For this, it can be useful to understand further the relation between the discrete and the continuous mean distance, since for instance, proving a constant factor relation between them would lead to subquadratic approximation algorithms for planar graphs.

Acknowledgements

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

[1] Abboud, A., Williams, V.V., Wang, J.R.: Approximation and fixed parameter subquadratic algorithms for radius and diameter in sparse graphs. In: Krauthgamer, R. (ed.) Proceedings of Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, pp. 377–391. SIAM (2016)

[2] Buckley, F.: Mean distance in line graphs. Congr. Numer. **32**, 153–162 (1976)

[3] Buckley, F., Superville, L.: Distance distributions and mean distance problems. In: Proceedings of 3rd Caribbean Conference on Combinatorics and Computing, pp. 67–76 (1981)

[4] Cabello, S.: Subquadratic algorithms for the diameter and the sum of pairwise distances in planar graphs. ACM Trans. Algorithms **15**(2), 21:1-21:38 (2018)

[5] Cabello, S., Knauer, C.: Algorithms for graphs of bounded treewidth via orthogonal range searching. Comput. Geom. **42**(9), 815–824 (2009)

[6] Dankelmann, P.: Average distance in weighted graphs. Discrete Math. **312**(1), 12–20 (2012)

[7] Dankelmann, P., Gutman, I., Mukwembi, S., Swart, H.C.: The edge-Wiener index of a graph. Discrete Math. **309**, 3452–3457 (2009)

[8] Doyle, J.K., Graver, J.E.: Mean distance in a graph. Discrete Math. **17**, 147–154 (1977)

[9] Doyle, J.K., Graver, J.E.: Mean distance for shapes. J. Graph Theory **6**(4), 453–471 (1982)

[10] Doyle, J.K., Graver, J.E.: A summary of results on mean distance in shapes. Environ. Plan. B Plan. Des. **9**, 177–179 (1982)

[11] Entringer, R.C., Jackson, D.E., Snyder, D.A.: Distance in graphs. Czech. Math. J. **26**, 283–296 (1976)

[12] Favaron, O., Kouider, M., Mahéo, M.: Edge-vulnerability and mean distance. Networks **19**(5), 493–504 (1989)

[13] Gutman, I.: A property of the Wiener number and its modifications. Indian J. Chem. **36**(A), 128–132 (1997)

[14] Henzinger, M.R., Klein, P., Rao, S., Subramanian, S.: Faster shortest-path algorithms for planar graphs. J. Comput. Syst. Sci. **55**(1), 3–23 (1997)

[15] Klamkin, M.S.: On the volume of a class of truncated prisms and some related centroid problems. Math. Mag. **41**(4), 175–181 (1968)

[16] Knor, M., Skrekovski, R., Tepeh, A.: Mathematical aspects of Wiener index. Ars Math. Contemp. **11**(2), 327–352 (2016)

[17] Kouider, M., Winkler, P.: Mean distance and minimum degree. J. Graph Theory **25**(1), 95–99 (1997)

[18] Lan, Y.F., Wang, Y.L.: An optimal algorithm for solving the 1-median problem on weighted 4-cactus graphs. Eur. J. Oper. Res. **122**(3), 602–610 (2000)

[19] March, L., Steadman, P.: The Geometry of Environment. Royal Institute of British Architects, London (1971)

[20] Merris, R.: An edge version of the matrix-tree theorem and the Wiener index. Linear Multilinear Algebra **25**(4), 291–296 (1989)

[21] Mohar, B.: Eigenvalues, diameter, and mean distance in graphs. Graphs Combin. **7**, 53–64 (1991)

[22] Ng, C.P., Teh, H.H.: On finite graphs of diameter 2. Nanta Math. **1**(72–75), 67 (1966)

[23] Nikolić, S., Trinajstić, N., Mihalić, Z.: The Wiener index: development and applications. Croat. Chem. Acta **68**, 105–129 (1995)

[24] Okabe, A., Boots, B., Sugihara, K.: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. Wiley Series in Probability and Mathematical Statistics, Wiley (1992)

[25] Otte, E., Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences. J. Inf. Sci. **28**, 441–453 (2002)

[26] Pettie, S., Ramachandran, V.: A shortest path algorithm for real-weighted undirected graphs. SIAM J. Comput. **34**(6), 1398–1431 (2005). https://doi.org/10.1137/S0097539702419650

[27] Plesník, J.: On the sum of all distances in a graph or a digraph. J. Graph Theory **8**(1), 1–21 (1984)

[28] Pletersek, P.Z.: The edge-Wiener index and the edge-hyper-Wiener index of phenylenes. Discrete Appl. Math. **255**, 326–333 (2019)

[29] Roditty, L., Williams, V.V.: Fast approximation algorithms for the diameter and radius of sparse graphs. In: Boneh, D., Roughgarden, T., Feigenbaum, J. (eds.) Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, pp. 515–524. ACM (2013)

[30] Rodríguez, J.A., Yebra, J.L.A.: Bounding the diameter and the mean distance of a graph from its eigenvalues: Laplacian versus adjacency matrix methods. Discrete Math. **196**(1–3), 267–275 (1999)

[31] Singh, Pradeep, Bhat, Vijay Kumar: Adjacency matrix and Wiener index of zero divisor graph $\Gamma(Z_n)$. J. Appl. Math. Comput. **66**(1), 717–732 (2021). https://doi.org/10.1007/s12190-020-01460-2

[32] Šoltés, Ľ: Transmission in graphs: a bound and vertex removing. Math. Slovaca **41**, 11–16 (1991)

[33] Thorup, M.: Undirected single-source shortest paths with positive integer weights in linear time. J. ACM **46**(3), 362–394 (1999)

[34] Weisstein, E.W.: Mean distance. From MathWorld—A Wolfram Web Resource. Last visited on 16/11/2020. https://mathworld.wolfram.com/MeanDistance.html

[35] Wiener, H.: Structural determination of paraffin boiling points. J. Am. Chem. Soc. **69**(1), 17–20 (1947)

[36] Zmazec, B., Zerovnik, J.: Estimating the traffic of weighted cactus networks in linear time. In: Proceedings of Ninth International Conference on Information Visualisation, pp. 1–6. SIAM (2005)

Delia Garijo and Alberto Márquez
Departamento de Matemática Aplicada I
Universidad de Sevilla
Avda. Reina Mercedes s/n
41012 Seville
Spain
e-mail: dgarijo@us.es;
       almar@us.es

Rodrigo I. Silveira
Departament de Matemàtiques
Universitat Politècnica de Catalunya
Jordi Girona 1-3
08034 Barcelona
Spain
e-mail: rodrigo.silveira@upc.edu

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.