Methodol Comput Appl Probab (2012) 14:299–334
DOI 10.1007/s11009-010-9204-9

An Investigation of New Graph Invariants Related to the Domination Number of Random Proximity Catch Digraphs

Elvan Ceyhan

Received: 19 November 2009 / Revised: 27 October 2010 /
Accepted: 28 October 2010 / Published online: 10 November 2010

Abstract Proximity catch digraphs (PCDs) are a special type of proximity graphs based on proximity maps which yield proximity regions. PCDs are defined using the relative allocation of points from two or more classes in a region of interest and have applications in various fields. We introduce some auxiliary tools for PCDs and graph invariants related to the domination number of the PCDs and investigate their probabilistic properties. We consider the cases in which the vertices of the PCDs come from uniform and non-uniform distributions in the region of interest. We also provide some of the newly defined proximity maps as illustrative examples.

Keywords Class cover catch digraph (CCCD) · Class cover problem · Covering set · Delaunay triangulation · Minimum dominating set · Proportional-edge PCD · Proximity graph · Random graph

AMS 2000 Subject Classification 60D05 · 05C80 · 05C20 · 60C05

1 Introduction

The proximity catch digraphs (PCDs) are a special type of proximity graphs which are based on proximity maps and are used in disciplines where shape and structure are important. Examples include computer vision (dot patterns), image analysis, pattern recognition (prototype selection), geography and cartography, visual perception, biology, and so on. *Proximity graphs* were introduced by Toussaint (1980), who called them *relative neighborhood graphs*. The notion of relative neighborhood graphs has been generalized in several directions and all of these graphs are now called proximity graphs.

E. Ceyhan (✉)
Department of Mathematics, Koç University, 34450, Sarıyer, Istanbul, Turkey
e-mail: elceyhan@ku.edu.tr

In recent years, a new classification and spatial pattern analysis approach which are based on the relative positions of the data points from various classes have been developed. Priebe et al. (2001) introduced the class cover catch digraphs (CCCDs) and gave the exact and the asymptotic distribution of the domination number of the CCCD based on two data sets, denoted $\mathcal{X}_n$ and $\mathcal{Y}_m$, both of which are random samples from uniform distribution on a compact interval in $\mathbb{R}$. DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a), Priebe et al. (2003b), and DeVinney and Priebe (2006) applied the concept in higher dimensions and demonstrated relatively good performance of CCCDs in classification. The employed methods involve data reduction (i.e., condensing) by using approximate minimum dominating sets as prototype sets, since finding the exact minimum dominating set is in general an NP-hard problem—in particular, for CCCDs (see DeVinney 2003). For the domination number of CCCDs based on one-dimensional data, a SLLN result is proved by DeVinney and Wierman (2003), and this result is extended by Wierman and Xiang (2008); furthermore, a CLT is also proved by Xiang and Wierman (2009). The asymptotic distribution of the domination number of CCCDs for non-uniform data in $\mathbb{R}$ is also calculated in a rather general setting (Ceyhan 2008). However, the exact and the asymptotic distribution of the domination number of the CCCDs are not analytically tractable in dimensions higher than one. Ceyhan (2005) extended the concept of CCCDs by introducing PCDs, which have have better properties (compared to CCCDs) in higher dimensions, such as the tractability of the distribution of the domination number. In particular, he introduced two new families of PCDs (namely, *proportional-edge* and *central similarity PCDs*). The distribution of the domination number of proportional-edge PCDs is derived and applied in spatial pattern analysis (Ceyhan and Priebe 2005, 2007; Ceyhan 2010). The distributions of the relative arc density of these PCD families are also derived and used for the same purpose (Ceyhan et al. 2006, 2007).

A general definition of proximity graphs is as follows: Let V be any finite or infinite set of points in $\mathbb{R}^d$. Each (unordered) pair of points $(p, q) \in V \times V$ is associated with a neighborhood $\mathfrak{N}(p, q) \subseteq \mathbb{R}^d$. Let $\mathfrak{P}$ be a property defined on $\mathfrak{N} = \{\mathfrak{N}(p, q) : (p, q) \in V \times V\}$. A *proximity* (or *neighborhood*) *graph* $G_{\mathfrak{N},\mathfrak{P}}(V, E)$ defined by the property $\mathfrak{P}$ is a graph with the set of vertices V and the set of edges E such that $pq \in E$ iff $\mathfrak{N}(p, q)$ satisfies property $\mathfrak{P}$. Examples of most commonly used proximity graphs are the Delaunay tessellation, the boundary of the convex hull, the Gabriel graph, relative neighborhood graph, Euclidean minimum spanning tree, and sphere of influence graph of a finite data set. See, e.g., Jaromczyk and Toussaint (1992). The relative allocation of the data points are used to construct a proximity digraph. A *digraph* is a directed graph, i.e., a graph with directed edges from one vertex to another based on a binary relation. Then the pair $(p, q) \in V \times V$ is an ordered pair and (p, q) is an arc (directed edge) denoted this way to reflect its difference from an edge. For example, the nearest neighbor (di)graph in Paterson and Yao (1992) is a proximity digraph. The nearest neighbor digraph, denoted $NND(V)$, has the vertex set V and (p, q) an arc iff $d(p, q) = \min_{v \in V \setminus \{p\}} d(p, v)$. That is, (p, q) is an arc of $NND(V)$ iff q is a nearest neighbor of p.

In this article, we introduce some graph invariants related to the domination number of the PCDs, provide a probabilistic investigation of the proximity maps and the associated regions. We present preliminaries and some PCD families in literature in Section 2 and the domination number of PCDs in Section 3, describe and

investigate the superset regions and Γ_1-regions in Section 4, introduce the concepts of the η- and κ-values for the PCDs in Sections 5 and 6. We present the Γ_k-regions for proximity maps in Section 7 and provide discussion and conclusions in Section 8.

2 Proximity Maps and Data-Random PCDs

2.1 Preliminaries

Let $(\Omega, \mathcal{M})$ be a measurable space and $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ and $\mathcal{Y}_m = \{Y_1, Y_2, \ldots, Y_m\}$ be two data sets from classes $\mathcal{X}$ and $\mathcal{Y}$ of Ω-valued random variables, respectively, whose joint probability density function (pdf) is $f_{X,Y}$. Let $d(\cdot,\cdot) : \Omega \times \Omega \to [0, \infty)$ be a distance function. The class cover problem for a target class, say $\mathcal{X}_n$, refers to finding a collection of neighborhoods, $N(X_i)$ around $X_i \in \mathcal{X}_n$ such that (i) $\mathcal{X}_n \subset \left(\bigcup_i N(X_i)\right)$ and (ii) $\mathcal{Y}_m \cap \left(\bigcup_i N(X_i)\right) = \emptyset$. A collection of neighborhoods satisfying both conditions is called a *class cover*. A cover satisfying condition (i) is a *proper cover* of class $\mathcal{X}$ while a collection satisfying condition (ii) is a *pure cover* relative to class $\mathcal{Y}$. From a practical point of view, for example for classification, of particular interest are the class covers satisfying both (i) and (ii) with the smallest collection of neighborhoods, i.e., minimum cardinality cover. This class cover problem is a generalization of the set cover problem of Garfinkel and Nemhauser (1972) that emerged in statistical pattern recognition and machine learning, where an edited or condensed set (i.e., prototype set) is selected from $\mathcal{X}_n$ (see, e.g., Devroye et al. 1996).

We construct the proximity regions using data sets from two classes. Given $\mathcal{Y}_m \subseteq \Omega$, the *proximity map* $N_{\mathcal{Y}}(\cdot) : \Omega \to \wp(\Omega)$ associates a *proximity region* $N_{\mathcal{Y}}(x) \subseteq \Omega$ with each point $x \in \Omega$, where $\wp(A)$ stands for the power set of A. The region $N_{\mathcal{Y}}(x)$ is defined in terms of the distance between x and $\mathcal{Y}_m$. More specifically, our proximity maps will be based on the relative position of points from class $\mathcal{X}$ with respect to the Delaunay tessellation of the points from class $\mathcal{Y}$. See Okabe et al. (2000) and Ceyhan (2009a) for more on Delaunay tessellations.

If $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ is a set of Ω-valued random variables, then $N_{\mathcal{Y}}(X_i)$ are random sets. If X_i are independent identically distributed, then so are the random sets $N_{\mathcal{Y}}(X_i)$. We define the data-random PCD D—associated with $N_{\mathcal{Y}}(\cdot)$—with vertex set $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ and arc set $\mathcal{A}$ by $(X_i, X_j) \in \mathcal{A} \iff X_j \in N_{\mathcal{Y}}(X_i)$. Since this relationship is not symmetric, a digraph is needed rather than a graph. The random digraph D depends on the (joint) distribution of the X_i and on the map $N_{\mathcal{Y}}(\cdot)$.

The PCDs might also be considered as a special case of *covering sets* of Tuza (1994) and *intersection digraphs* of Sen et al. (1989). This data random proximity digraph is a *vertex-random proximity digraph* which is not of standard type. The randomness of the PCDs lies in the fact that the vertices are random with joint pdf $f_{X,Y}$, but arcs (X_i, X_j) are deterministic functions of the random variable X_i and the set $N_{\mathcal{Y}}(X_i)$. For example, the CCCD of Priebe et al. (2001) can be viewed as an example of PCD with $N_{\mathcal{Y}}(x) = B(x, r(x))$, where $r(x) := \min_{\mathsf{y} \in \mathcal{Y}_m} d(x, \mathsf{y})$. The CCCD is the digraph of order n with vertex set $\mathcal{X}_n$ and an arc from X_i to X_j iff $X_j \in B(X_i, r(X_i))$. That is, there is an arc from X_i to X_j iff there exists an open ball centered at X_i which is

"pure" (or contains no elements) of $\mathcal{Y}_m$, and simultaneously contains (or "catches") point X_j.

2.2 Spherical and Arc-Slice PCDs

Let $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_m\} \subset \mathbb{R}^d$. For $\Omega = \mathbb{R}^d$, the proximity map associated with CCCD is defined as the open ball $N_S(x) := B(x, r(x))$ for all $x \in \mathbb{R}^d \setminus \mathcal{Y}_m$ (hence the name *spherical proximity map* and the notation N_S) and for $x \in \mathcal{Y}_m$, define $N_S(x) = \{x\}$. Furthermore, dependence on $\mathcal{Y}_m$ is through $r(x)$. Hence, for $d = 1$, $N_S(x)$ is based on the intervals $I_i = \left(\mathsf{y}_{(i):m}, \mathsf{y}_{(i+1):m}\right)$ for $i = 0, 1, 2, \ldots, m$ with $\mathsf{y}_{0:m} = -\infty$ and $\mathsf{y}_{(m+1):m} = \infty$ where $\mathsf{y}_{i:m}$ is the i^{th} order statistic in $\mathcal{Y}_m$. This intervalization can be viewed as a tessellation, since it partitions the convex hull of $\mathcal{Y}_m$, denoted $\mathcal{C}_H(\mathcal{Y}_m)$. For $d > 1$, a natural tessellation that partitions $\mathcal{C}_H(\mathcal{Y}_m)$ is the Delaunay tessellation (see Okabe et al. 2000; Ceyhan 2009a). Let $\mathcal{T}_i$ be the i^{th} Delaunay cell in the Delaunay tessellation of $\mathcal{Y}_m$ for $i = 1, 2, \ldots, J$. In $\mathbb{R}$, we implicitly use the cell that contains x to define the proximity region.

The spherical proximity map $N_S(x)$ is well-defined for all $x \in \mathbb{R}^d$ provided that $\mathcal{Y}_m \neq \emptyset$. The spherical proximity map—with applications in classification—is investigated in DeVinney et al. (2002), DeVinney and Wierman (2003), Marchette and Priebe (2003), Priebe et al. (2003a, b), and DeVinney and Priebe (2006). However, for $d > 1$, finding the minimum dominating set of the CCCD (i.e., the PCD associated with $N_S(\cdot)$) is an NP-hard problem and the distribution of the domination number is not analytically tractable (Ceyhan 2005). Note that for $d = 1$, such problems do not exist. After a slight modification, the spherical proximity maps give rise to *arc-slice proximity maps* which are defined as $N_{AS}(x) := B(x, r(x)) \cap \mathcal{T}_i$ for $x \in \mathcal{T}_i$ (i.e., the arc-slice proximity region is the spherical proximity region restricted to the Delaunay cell that contains x). Notice that in $\mathbb{R}$, the proximity maps N_S and N_{AS} are equivalent (i.e., they yield the same proximity region for each x). However, when $x \notin \mathcal{C}_H(\mathcal{Y}_m)$ for $d > 1$ (i.e., x is not in any of the Delaunay cells based on $\mathcal{Y}_m$), $N_{AS}(x)$ is not defined.

2.3 Vertex and Edge Regions

In $\mathbb{R}$, the spherical proximity maps are defined as open intervals with one of the endpoints being in $\mathcal{Y}_m$. In particular, for $x \in I_i = (\mathsf{y}_{i:m}, \mathsf{y}_{(i+1):m})$ for $i = 1, 2, \ldots, m$, $N_S(x) = (\mathsf{y}_{i:m}, \mathsf{y}_{i:m} + 2r(x))$ and $r(x) = d(x, \mathsf{y}_{i:m})$ for all $x \in (\mathsf{y}_{i:m}, (\mathsf{y}_{(i+1):m} + \mathsf{y}_{i:m})/2)$ and $N_S(x) = (\mathsf{y}_{(i+1):m} - 2r(x), \mathsf{y}_{(i+1):m})$ and $r(x) = d(x, \mathsf{y}_{(i+1):m})$ for all $x \in ((\mathsf{y}_{(i+1):m} + \mathsf{y}_{i:m})/2, \mathsf{y}_{(i+1):m})$. Hence there are two subintervals in I_i each touching an end point and the midpoint of the interval and $N_S(x)$ depends on which of these regions x lies in.

In $\mathbb{R}^d$ with $d > 1$, intervals become Delaunay cells and our proximity maps are based on the Delaunay cell $\mathcal{T}_i$ that contains x. The region $N_{\mathcal{Y}}(x)$ will also depend on the location of x in $\mathcal{T}_i$ with respect to the vertices or faces (i.e., edges in $\mathbb{R}^2$) of $\mathcal{T}_i$. Hence for $N_{\mathcal{Y}}(x)$ to be well-defined, the vertex or face of $\mathcal{T}_i$ associated with x should be uniquely determined. This gives rise to two new concepts, namely, *vertex regions* and *face regions* (*edge regions* in $\mathbb{R}^2$).

Definition 2.1 (Vertex Regions) Let $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}_3) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ be the triangle (including the interior) with vertices $\mathcal{Y}_3$. The connected regions that partition $T(\mathcal{Y}_3)$, (in the sense that the pairwise intersections of the regions have zero $\mathbb{R}^2$-Lebesgue measure) such that each region has one and only one vertex of $T(\mathcal{Y}_3)$ on its boundary are called *vertex regions*.

In fact, we can construct the vertex regions using a point $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, in particular, M being a *center* of $T(\mathcal{Y}_3)$ by joining M to the edges with *(straight) line segments*. We call such regions as *M-vertex regions* and denote the vertex region associated with vertex y as $R_M(\mathsf{y})$ for $\mathsf{y} \in \mathcal{Y}_3$. Vertex regions with circumcenter, incenter, and center of mass are investigated in Ceyhan (2009a). See Fig. 1 (left) for M-vertex regions defined by using the extensions of the line segments joining y to M for each $\mathsf{y} \in \mathcal{Y}_3$ with M being the center of mass, M_C. Let e_i be the edge opposite vertex y_i and M_i be the midpoint of edge e_i for $i = 1, 2, 3$. The lines joining y_i to M_i pass through M_C and are called *median lines*.

Note that the endpoints of an interval constitute its boundary, but for a triangle, the boundary consists of the vertices and edges, which suggests the concept of *edge regions*.

Definition 2.2 (Edge Regions) The connected regions that partition the triangle, $T(\mathcal{Y}_3)$, in such a way that each region has one and only one edge of $T(\mathcal{Y}_3)$ on its boundary, are called *edge regions*.

In fact, we can construct the edge regions starting with any point M in $T(\mathcal{Y}_3)^o$, where A^o stands for the interior of A, (e.g., M can be a center of the triangle (Kimberling 2008)) by joining the point M to the vertices by straight lines. We call such regions *M-edge regions* and denote the region for edge e as $R_M(e)$ for $e \in \{e_1, e_2, e_3\}$. See Fig. 1 (right) with $M = M_C$. Edge regions for incenter, center of mass, and orthocenter are investigated in Ceyhan (2009a).

In $\mathbb{R}^d$ with $d > 2$, let $\mathcal{Y}_{d+1} = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_{d+1}\}$ be $d+1$ points that do not lie on the same $(d-1)$-dimensional hyperplane. Denote the simplex formed by these $d+1$ points as $\mathcal{S}(\mathcal{Y}_{d+1})$. A simplex is the simplest polytope in $\mathbb{R}^d$ having $d+1$ vertices, $d(d+1)/2$ edges, and $d+1$ faces of dimension $(d-1)$. The connected regions that partition $\mathcal{S}(\mathcal{Y}_{d+1})$, (in the sense that the pairwise intersections of the regions have

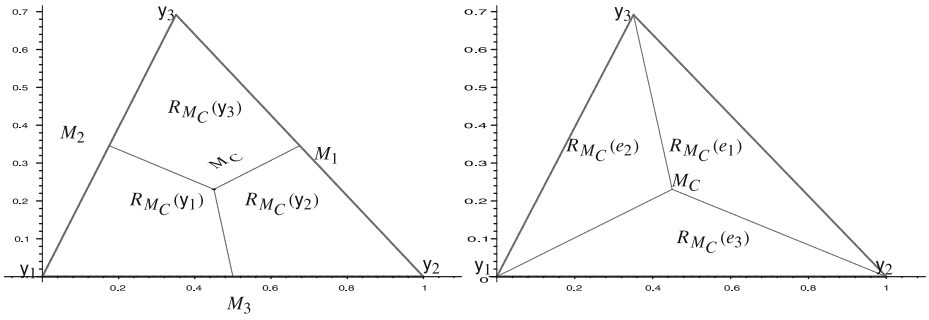

Fig. 1 An illustration of M_C-vertex regions, $R_{M_C}(\mathsf{y}_i)$ (*left*) and M_C-edge regions, $R_{M_C}(e_i)$ (*right*) with median lines for $i = 1, 2, 3$

zero $\mathbb{R}^{d-1}$-Lebesgue measure) such that each region has one and only one vertex of $\mathfrak{S}(\mathcal{Y}_{d+1})$ on its boundary are the *vertex regions*. Similarly, the connected regions that partition the simplex, $\mathfrak{S}(\mathcal{Y}_{d+1})$, in such a way that each region has one and only one face of $\mathfrak{S}(\mathcal{Y}_{d+1})$ on its boundary are the *face regions*.

2.4 Proportional-Edge Proximity Maps

Let $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_m\}$ be m points in general position in $\mathbb{R}^d$ and $\mathcal{T}_i$ be the i^{th} Delaunay cell for $i = 1, 2, \ldots, J$. Let $\mathcal{X}_n$ be a random sample from a distribution F in $\mathbb{R}^d$ with support $\mathcal{S}(F) \subseteq \mathcal{C}_H(\mathcal{Y}_m)$. For illustrative purposes, suppose $d = 2$, then a Delaunay tessellation is a triangulation, provided that no more than three points in $\mathcal{Y}_m$ are cocircular. Furthermore, for simplicity, let $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}_3) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ be the triangle with vertices $\mathcal{Y}_3$. Let $\mathcal{X}_n$ be a random sample from F with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$.

For $r \in [1, \infty]$, define $N_{PE}^r(\cdot, M) := N(\cdot, M; r, \mathcal{Y}_3)$ to be the *proportional-edge proximity map* with M-vertex regions as follows (see also Fig. 2 with $M = M_C$ and $r = 2$). For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $v(x) \in \mathcal{Y}_3$ be the vertex whose region contains x; i.e., $x \in R_M(v(x))$. If x falls on the boundary of two M-vertex regions, we assign $v(x)$ arbitrarily. Let $e(x)$ be the edge of $T(\mathcal{Y}_3)$ opposite $v(x)$. Let $\ell(v(x), x)$ be the line parallel to $e(x)$ through x. Let $d(v(x), \ell(v(x), x))$ be the Euclidean distance from $v(x)$ to $\ell(v(x), x)$. Then for $r \geq 1$

$$N_{PE}^r(x, M) := \{z \in R_M(v(x)) : d(v(x), \ell(v(x), z)) \leq r\, d(v(x), \ell(v(x), x))\}.$$

Notice that $\ell(v(x), x)$ divides the edges of $T_r(x)$ (other than $\ell_r(v(x), x)$) proportionally with the factor r. Hence the name *proportional edge proximity region*.

Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M)$. Furthermore, $\lim_{r\to\infty} N_{PE}^r(x, M) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, so we define $N_{PE}^\infty(x, M) = T(\mathcal{Y}_3)$ for all such x. For $x \in \mathcal{Y}_3$, we define $N_{PE}^r(x, M) = \{x\}$ for all $r \in [1, \infty]$. Furthermore, $X_i \overset{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional pdf f exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$, implies that the special case in the construction of N_{PE}^r—X falls on the boundary of two vertex regions—occurs with probability zero. Note that for such an F, $N_{PE}^r(X)$ is a triangle a.s. Of particular interest is $N_{PE}^r(x, M)$ with any

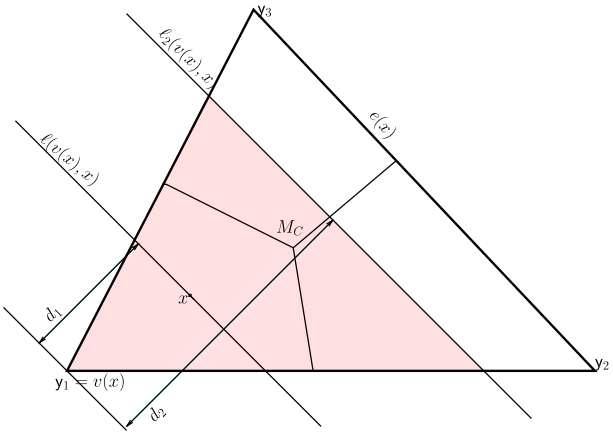

Fig. 2 Construction of proportional edge proximity region, $N_{PE}^2(x, M_C)$ (*shaded region*) where $d_1 = d(v(x), \ell(v(x), x))$ and $d_2 = d(v(x), \ell_2(v(x), x)) = 2\, d(v(x), \ell(v(x), x))$

M and $r \in \{\sqrt{2}, 2\}$. For $r = \sqrt{2}$, $\ell(v(x), x)$ divides $T_{\sqrt{2}}(x)$ into two regions of equal area, hence $N_{PE}^{\sqrt{2}}(x, M)$ is also referred to as *double-area proximity region*. For $r = 2$, $\ell(v(x), x)$ divides the edges of $T_2(x)$—other than $\ell_r(v(x), x)$—into two segments of equal length, hence $N_{PE}^{2}(x, M)$ is also referred to as *double-edge proximity region*.

2.4.1 Extension of N_{PE}^r to Higher Dimensions

The extension of N_{PE}^r to $\mathbb{R}^d$ for $d > 2$ is straightforward. The extension with $M = M_C$ is provided here; the extension for general M is similar. For $r \in [1, \infty]$, define the proximity map as follows. Given a point x in $\mathfrak{S}(\mathcal{Y}_{d+1})$, let $v := \text{argmin}_{\mathsf{y} \in \mathcal{Y}_{d+1}} \mathrm{V}(Q_{\mathsf{y}}(x))$ where $Q_{\mathsf{y}}(x)$ is the polytope with vertices being the $d(d+1)/2$ midpoints of the edges, the vertex v and x and $\mathrm{V}(\cdot)$ is the d-dimensional volume function. That is, the vertex region for vertex v is the polytope with vertices v, center of mass, and the midpoints of the edges adjacent to v. Let $v(x)$ be the vertex in whose region x falls. If x falls on the boundary of two vertex regions, $v(x)$ is assigned arbitrarily. Let $\varphi(x)$ be the face opposite to vertex $v(x)$, and $\Upsilon(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ which contains x. Let $d(v(x), \Upsilon(v(x), x))$ be the Euclidean distance from $v(x)$ to $\Upsilon(v(x), x)$. For $r \in [1, \infty)$, let $\Upsilon_r(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ such that

$$d(v(x), \Upsilon_r(v(x), x)) = r\, d(v(x), \Upsilon(v(x), x)) \text{ and}$$
$$d(\Upsilon(v(x), x), \Upsilon_r(v(x), x)) < d(v(x), \Upsilon_r(v(x), x)).$$

Let $\mathfrak{S}_r(x)$ be the polytope similar to and with the same orientation as $\mathfrak{S}(\mathcal{Y}_{d+1})$ having $v(x)$ as a vertex and $\Upsilon_r(v(x), x)$ as the opposite face. Then the proximity region is defined as $N_{PE}^r(x, M_C) := \mathfrak{S}_r(x) \cap \mathfrak{S}(\mathcal{Y}_{d+1})$. Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M_C)$. The special cases in the construction of $N_{PE}^r(x, M_C)$ can be handled as in the two-dimensional case.

3 Domination Number of the PCDs and Related Concepts

In a digraph $D = (\mathcal{V}, \mathcal{A})$, a vertex $v \in \mathcal{V}$ *dominates* itself and all vertices of the form $\{u : (v, u) \in \mathcal{A}\}$. A *dominating set* S_D for the digraph D is a subset of $\mathcal{V}$ such that each vertex $v \in \mathcal{V}$ is dominated by a vertex in S_D. A *minimum dominating set* S_D^* is a dominating set of minimum cardinality and the *domination number* $\gamma(D)$ is defined as $\gamma(D) := |S_D^*|$ (see, e.g., Lee 1998) where $|\cdot|$ denotes the set cardinality function.

For $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$ the domination number of the associated data-random PCD, denoted $\gamma_n(N)$, is the minimum number of points that dominate all points in $\mathcal{X}_n$.

3.1 Distribution of the Domination Number of the PCDs in Literature

Let $\gamma_n(N_S)$ be the domination number of the PCD based on N_S (i.e., CCCD) for uniform data in $(\mathsf{y}_1, \mathsf{y}_2) \subset \mathbb{R}$ with $-\infty < \mathsf{y}_1 < \mathsf{y}_2 < \infty$. Then the exact distribution of $\gamma_n(N_S)$ is

$$\gamma_n(N_S) = \begin{cases} 1 \text{ w.p. } 5/9 + (16/9)\, 4^{-n}, \\ 2 \text{ w.p. } 4/9 - (16/9)\, 4^{-n}, \end{cases} \text{ for all } n \geq 1, \tag{1}$$

where w.p. stands for "with probability". See Priebe et al. (2001) for the derivation. This result is extended for non-uniform data in $\mathbb{R}$ in Ceyhan (2008).

Let $\gamma_n(r)$ be the domination number of the PCD based on $N^r_{PE}(\cdot, M_C)$ for uniform data in $T(\mathcal{Y}_3) \subset \mathbb{R}^2$ where $T(\mathcal{Y}_3) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$. For $r = 3/2$, $\lim_{n\to\infty} \gamma_n(r) > 1$ a.s. Furthermore,

$$\lim_{n\to\infty} \gamma_n(3/2) = \begin{cases} 2 & \text{wp} \approx 0.7413, \\ 3 & \text{wp} \approx 0.2487. \end{cases}$$

See Ceyhan and Priebe (2005) for the derivation. This result is further extended for the entire range of the parameter r and the center M by Ceyhan and Priebe (2007).

The main strategy used in these articles is first showing that the distribution of the domination number for CCCD in $\mathbb{R}$ and for the PCD based on N^r_{PE} in $\mathbb{R}^d$ is independent of the geometry of the support (i.e., geometry invariant) for uniform data. This enables one to first compute the exact and asymptotic distributions for data in one Delaunay cell and then extend the result to multiple Delaunay cell case easily. Below we provide the details of the transformations that yield the geometry invariance result.

In $\mathbb{R}^d$ with $d > 1$, the exact and asymptotic distribution of the domination number of the PCDs based on N_S and N_{AS} are still open problems (Ceyhan 2005). However, the asymptotic distribution of the domination number of the PCD based on N^r_{PE} and uniform data in $\mathbb{R}^d$ with $d > 2$ can be derived similarly as in the two-dimensional case.

3.2 Transformations Preserving Uniformity of Data on Intervals and Triangles

The proximity regions (hence the corresponding PCDs) are based on the Delaunay tessellation of $\mathcal{Y}_m$, which partitions $\mathcal{C}_H(\mathcal{Y}_m)$. Suppose the set $\mathcal{X}_n$ is a set of iid uniform random variables on the convex hull of $\mathcal{Y}_m$; i.e., a random sample from $\mathcal{U}(\mathcal{C}_H(\mathcal{Y}_m))$. In particular, conditional on $|\mathcal{X}_n \cap \mathcal{T}_i| > 0$ being fixed, $\mathcal{X}_n \cap \mathcal{T}_i$ will also be a set of iid uniform random variables on $\mathcal{T}_i$ for $i \in \{1, 2, \ldots, J\}$. Reducing the cell $\mathcal{T}_i$ as much as possible while preserving uniformity and the probabilities related to PCDs will simplify the notation and calculations.

In $\mathbb{R}$ the Delaunay cells $\mathcal{T}_i$ are intervals, I_i. Let $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\} \subset \mathbb{R}$ be two non-concurrent points. Without loss of generality, assume $\mathsf{y}_1 < \mathsf{y}_2$, and let $X_i \overset{iid}{\sim} \mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$ for $i = 1, 2, \ldots, n$. The pdf of $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$ is $f(u) = \frac{1}{\mathsf{y}_2 - \mathsf{y}_1} \mathbf{I}(u \in (\mathsf{y}_1, \mathsf{y}_2))$. The interval $(\mathsf{y}_1, \mathsf{y}_2)$ can be mapped to the unit interval by $\phi(x) = (x - \mathsf{y}_1)/(\mathsf{y}_2 - \mathsf{y}_1)$. A quick investigation shows that the random variables $X_i \overset{iid}{\sim} \mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$ transformed along with $(\mathsf{y}_1, \mathsf{y}_2)$ in the described fashion by ϕ satisfy $\phi(X_i) \overset{iid}{\sim} \mathcal{U}(0, 1)$. So, without loss of generality, we can assume the interval $(\mathsf{y}_1, \mathsf{y}_2)$ to be the unit interval $(0, 1)$ for uniform data.

In $\mathbb{R}^2$ the Delaunay cells $\mathcal{T}_i$ are triangles. Let $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\} \subset \mathbb{R}^2$ be three non-collinear points and $T(\mathcal{Y}_3) = \mathcal{T}_1$ be the triangle with vertices $\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3$. Let $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ for $i = 1, 2, \ldots, n$. The pdf of $\mathcal{U}(T(\mathcal{Y}_3))$ is

$$f(u) = \frac{1}{A(T(\mathcal{Y}_3))} \mathbf{I}(u \in T(\mathcal{Y}_3)),$$

where $A(\cdot)$ is the area function. The triangle $T(\mathcal{Y}_3)$ can be carried into the first quadrant by a composition of (some of the) transformations of scaling, translation, rotation, and reflection in such a way that the largest edge has unit length and lies on the x-axis, and the x-coordinate of the vertex nonadjacent to largest edge is less than $1/2$. We call the resultant triangle the *basic triangle* and denote it as T_b, so we have $T_b = \big((0,0),(1,0),(c_1,c_2)\big)$ with $0 < c_1 \le 1/2$, and $c_2 > 0$ and $(1-c_1)^2 + c_2^2 \le 1$. The transformation from any triangle to T_b is denoted by ϕ_b. Notice that if $T(\mathcal{Y}_3)$ is transformed into T_b, then $T(\mathcal{Y}_3)$ is similar to T_b and $\phi_b(T(\mathcal{Y}_3)) = T_b$. Thus the random variables $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ transformed along with $T(\mathcal{Y}_3)$ in the described fashion by ϕ_b satisfy $\phi_b(X_i) \overset{iid}{\sim} \mathcal{U}(T_b)$. So, without loss of generality, we can assume $T(\mathcal{Y}_3)$ to be T_b for uniform data for any proximity map $N(\cdot)$.

There are many transformations that preserve uniformity of the data on a triangle, but not all of them preserve the similarity of the triangles. That is, a uniformity preserving transformation when applied to similar triangles might yield non-similar triangles. Hence we only describe the transformation that maps T_b to the standard equilateral triangle, $T_e = T((0,0),(1,0),(1/2,\sqrt{3}/2))$ for exploiting the symmetry in calculations using T_e.

Let $\phi_e : (x,y) \to (u,v)$, where $u(x,y) = x + \frac{1-2c_1}{2c_2}\, y$ and $v(x,y) = \frac{\sqrt{3}}{2c_2}\, y$. Then in T_b, the vertices $\mathsf{y}_1 = (0,0)$, $\mathsf{y}_2 = (1,0)$, and $\mathsf{y}_3 = (c_1,c_2)$ are mapped to $(0,0)$, $(1,0)$, and $\big(1/2, \sqrt{3}/2\big)$, respectively. See also Fig. 3.

Note that the inverse transformation is $\phi_e^{-1}(u,v) = \big(x(u,v), y(u,v)\big)$ where $x(u,v) = u - \frac{(1-2c_1)}{\sqrt{3}}\, v$ and $y(u,v) = \frac{2c_2}{\sqrt{3}}\, u$. Then the Jacobian is $J(x,y) = 2c_2/\sqrt{3}$ (Ceyhan 2009b). So

$$f_{U,V}(u,v) = f_{X,Y}(\phi_e^{-1}(u,v))\,|J| = \frac{4}{\sqrt{3}}\,\mathbf{I}\big((u,v) \in T_e\big).$$

Hence uniformity is preserved.

Remark 3.1 The probability of an event for uniform data in $T(\mathcal{Y}_3)$ involves the ratio of the area of the region associated with the event to $A(T(\mathcal{Y}_3))$. If such ratios are not preserved under ϕ_e, then the probability content for N depends on the geometry of

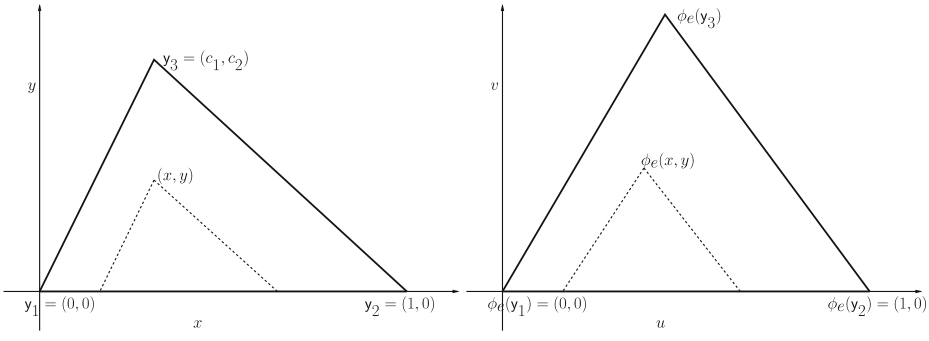

Fig. 3 The description of $\phi_e(x,y)$ for $(x,y) \in T_b$ (*left*) and the equilateral triangle $\phi_e(T_b) = T_e$ (*right*)

$T(\mathcal{Y}_3)$. In particular, the probability content for uniform data for N_S and N_{AS} depend on the geometry of the triangle, hence is not geometry invariant (Ceyhan 2009a). For example, $P(X \in N_S(Y))$ and $P(X \in N_{AS}(Y, M))$ depends on (c_1, c_2), hence one has to do the computations for all of these triangles (but unfortunately there are uncountably many such triangles. But for a proximity region $N(\cdot)$, if the probability content for uniform data is preserved after ϕ_e is applied on the data, then we say the PCD is *geometry invariant* for uniform data. Hence, without loss of generality, we can assume $T(\mathcal{Y}_3)$ to be T_e for uniform data for such proximity maps.

In $\mathbb{R}^d$ with $d > 2$, any Delaunay cell (which is now a d-dimensional polytope with $d+1$ vertices) can be mapped into the (standard) regular polytope in $\mathbb{R}^d$, using transformations similar to the ones for the two-dimensional case. For example in $\mathbb{R}^3$, a Delaunay cell is a tetrahedron. By a combination of (some of) the transformations of scaling, translation, rotation, and reflection, this tetrahedron can be mapped to a tetrahedron (called the *basic tetrahedron*) with vertices $(0,0,0)$, $(1,0,0)$, $(c_1, c_2, 0)$, and (d_1, d_2, d_3) in such a way that the largest face lies in the xy-plane, and the vertex (d_1, d_2, d_3) is the closest vertex (other than $(0,0,0)$) to the z-axis. Then it is also possible to map the basic tetrahedron to the standard regular tetrahedron with vertices $(0,0,0)$, $(1,0,0)$, $\left(1/2, \sqrt{3}/2, 0\right)$, and $\left(1/2, \sqrt{3}/6, \sqrt{6}/3\right)$ preserving the uniformity of the data in the basic tetrahedron. In fact, for this purpose, we can use the transformation $\varphi(x, y, z) = (u, v, w)$ where $u(x, y, z) = x + \frac{1-2c_1}{2c_2} y + \frac{c_2(1-2d_1)-d_2(1-2c_1)}{2c_2 d_3} z$, $v(x, y, z) = \frac{\sqrt{3}}{2c_2} y + \frac{\sqrt{3}(c_2-3d_2)}{6c_2 d_3} z$, and $w(x, y, x) = \frac{\sqrt{6}}{3d_3} z$. Furthermore, the Jacobian for this transformation can be shown to be $|J| = \sqrt{2} c_2 d_3$.

4 Superset Regions and Γ_1-Regions

First consider the case of data on $\mathbb{R}$. For $x \in I_i$, the spherical proximity region $N_S(x) = I_i$ iff $x = \left(\mathsf{y}_{i:m} + \mathsf{y}_{(i+1):m}\right)/2$. We define an associated region for such points in the general context.

Definition 4.1 (Superset Region) The *superset region* for any proximity map $N(\cdot)$ in Ω is defined to be $\mathscr{R}_S(N) := \left\{x \in \Omega : N(x) = \Omega\right\}$. When X is Ω-valued random variable, then $X \in \mathscr{R}_S(N)$ if $N(X) = \Omega$ a.s.

For example, for $\Omega = I_i \subsetneq \mathbb{R}$ with $i = 1, 2, \ldots, (m-1)$, $\mathscr{R}_S(N_S) = \{x \in I_i : N_S(x) = I_i\} = \left\{\left(\mathsf{y}_{i:m} + \mathsf{y}_{(i+1):m}\right)/2\right\}$; and for $i = 0, m$ (i.e., $\Omega = I_0$ or $\Omega = I_m$), we have $\mathscr{R}_S(N_S) = \emptyset$ since $N_S(x) \subsetneq I_i$ for all $x \in I_i$. More generally for $\Omega = \mathcal{T}_i \subsetneq \mathbb{R}^d$ (i.e., Ω is the i^{th} Delaunay cell), we have $\mathscr{R}_S(N_S) = \{x \in \mathcal{T}_i : N_S(x) = \mathcal{T}_i\}$. Note that for $x \in I_i$, it follows that $\lambda(N_S(x)) \leq \lambda(I_i)$ and moreover $\lambda(N_S(x)) = \lambda(I_i)$ iff $x \in \mathscr{R}_S(N_S)$ where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}$ (also called as $\mathbb{R}$-Lebesgue measure). So the proximity region of a point in $\mathscr{R}_S(N_S)$ has the largest $\mathbb{R}$-Lebesgue measure. Note that for $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_m\} \subset \mathbb{R}$ (i.e., $\Omega = \mathbb{R}$), and $\mathcal{X}_n \cap I_k$ is not empty for at least two of $k = 1, 2, \ldots, (m-1)$, we have $\mathscr{R}_S(N_S) = \emptyset$, since $N_S(x) \subseteq I_i$ for all $x \in I_i$, so $N_S(x) \subsetneq \mathbb{R}$ for all $x \in \mathbb{R}$. Note also that for a given realization of $\mathcal{Y}_m$, $\mathscr{R}_S(N_S)$ is not a random set, but $\mathbf{I}(X \in \mathscr{R}_S(N_S))$ is a random variable.

Definition 4.2 (Γ_1-Region) Let $(\Omega, \mathcal{M})$ be a measurable space and consider the proximity map $N : \Omega \to \wp(\Omega)$. For any set $B \subseteq \Omega$, the Γ_1-*region* of B associated with $N(\cdot)$, is defined to be the region $\Gamma_1(B, N) := \{z \in \Omega : B \subseteq N(z)\}$.

For $x \in \Omega$, we denote $\Gamma_1\big(\{x\}, N\big)$ as $\Gamma_1(x, N)$. Note that Γ_1-region is based on the proximity region $N(\cdot)$. If $\mathcal{X}_n$ is a set of Ω-valued random variables, then $\Gamma_1(X_i, N)$, $i = 1, 2, \ldots, n$ are random sets. If the X_i are independent and identically distributed, then so are the random sets $\Gamma_1(X_i, N)$. Additionally, $\Gamma_1(\mathcal{X}_n, N)$ is also a random set. Note that, $\gamma_n(N) = 1$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N) \neq \emptyset$. Hence the name Γ_1-*region*.

4.1 General Results on Superset and Γ_1-Regions

Suppose μ is a measure on Ω. Following are some general results about superset regions and Γ_1-regions.

Proposition 4.3 *For any proximity map N and set $B \subseteq \Omega$, we have $\mathscr{R}_S(N) \subseteq \Gamma_1(B, N)$.*

Proof For $x \in \mathscr{R}_S(N)$, it follows that $N(x) = \Omega$, so $B \subseteq N(x)$, since $B \subseteq \Omega$. Then $x \in \Gamma_1(B, N)$, hence $\mathscr{R}_S(N) \subseteq \Gamma_1(B, N)$. □

In fact, if $B = \Omega$, then $\Gamma_1(\Omega, N) = \mathscr{R}_S(N)$.

Lemma 4.4 *For any proximity map N and set $B \subseteq \Omega$, we have $\Gamma_1(B, N) = \bigcap_{x \in B} \Gamma_1(x, N)$.*

Proof Given a proximity map N and subset $B \subseteq \Omega$, $y \in \Gamma_1(B, N)$ iff $B \subseteq N(y)$ iff $x \in N(y)$ for all $x \in B$ iff $y \in \Gamma_1(x, N)$ for all $x \in B$ iff $y \in \bigcap_{x \in B} \Gamma_1(x, N)$. Hence the result follows. □

Corollary 4.5 *For any proximity map N and a realization $\mathcal{X}_n = \{x_1, x_2, \ldots, x_n\}$ from a distribution with support in Ω, we have $\Gamma_1(\mathcal{X}_n, N) = \bigcap_{i=1}^{n} \Gamma_1(x_i, N)$.*

Lemma 4.6 *Given a sequence of Ω-valued random variables $X_1, X_2, \ldots$, let $\mathcal{X}(n) := \mathcal{X}(n-1) \cup \{X_n\}$ for $n = 1, 2, \ldots$ with $\mathcal{X}(0) := \emptyset$. Then $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing in n in the sense that $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$.*

Proof Given a particular type of proximity map N and a data set $\mathcal{X}(n) = \{X_1, X_2, \ldots, X_n\}$, by Lemma 4.4, it follows that $\Gamma_1(\mathcal{X}(n), N) = \bigcap_{i=1}^{n} \Gamma_1(X_i, N)$ and by definition, we have $\mathcal{X}(n+1) = \mathcal{X}(n) \cup \{X_{n+1}\}$. So,

$$\begin{aligned}
\Gamma_1(\mathcal{X}(n+1), N) &= \bigcap_{i=1}^{n+1} \Gamma_1(X_i, N) \\
&= \left[\bigcap_{i=1}^{n} \Gamma_1(X_i, N)\right] \bigcap \Gamma_1(X_{n+1}, N) \\
&= \Gamma_1(\mathcal{X}(n), N) \cap \Gamma_1(X_{n+1}, N) \subseteq \Gamma_1(\mathcal{X}(n), N).
\end{aligned}$$

Thus we have shown that $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing in n; i.e., $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$. □

Theorem 4.7 *Given a sequence of random variables $X_1, X_2, \ldots$ which are identically distributed on Ω, let $\mathcal{X}(n) := \mathcal{X}(n-1) \cup \{X_n\}$ for $n = 1, 2, 3, \ldots$ and $\mathcal{X}(0) := \emptyset$. Then $\Gamma_1(\mathcal{X}(n), N) \downarrow \mathscr{R}_S(N)$, as $n \to \infty$ a.s. in the sense that $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$ and $\mu(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)) \downarrow 0$ a.s.*

Proof By Lemma 4.6, we have $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$. By monotone sequential continuity from above (Billingsley 1995), the sequence $\{\Gamma_1(\mathcal{X}(n), N)\}_{n=1}^{\infty}$ has a limit

$$G_1 := \bigcap_{j=1}^{\infty} \Gamma_1(\mathcal{X}(j), N) = \lim_{m\to\infty} \bigcap_{j=1}^{m} \Gamma_1(\mathcal{X}(j), N) = \lim_{m\to\infty} \Gamma_1(\mathcal{X}(m), N)$$

$$= \lim_{m\to\infty} \bigcap_{i=1}^{m} \Gamma_1(X_i, N) = \bigcap_{i=1}^{\infty} \Gamma_1(X_i, N). \tag{2}$$

We claim that $G_1 = \mathscr{R}_S(N)$ a.s.

Suppose $\mathscr{R}_S(N) \subsetneq \Omega$, since if $\mathscr{R}_S(N) = \Omega$, then $N(x) = \Omega$ for all $x \in \Omega$, so $\Gamma_1(\mathcal{X}(n), N) = \Omega$ for all x and $n > 1$ hence the result would follow trivially. Since $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}(n), N)$ for all n, it follows that $\mathscr{R}_S(N) \subseteq G_1$. Let $\varepsilon > 0$. Then

$$P\left(\sup_{k\geq n} \mu\left(\bigcap_{i=1}^{k} \Gamma_1(X_i, N) \setminus \mathscr{R}_S(N)\right) \leq \varepsilon\right) = P\left(\mu\left(\bigcap_{i=1}^{n} \Gamma_1(X_i, N) \setminus \mathscr{R}_S(N)\right) \leq \varepsilon\right)$$

$$= P\left(\mu\left(\Gamma_1(\mathcal{X}(i), N) \setminus \mathscr{R}_S(N)\right) \leq \varepsilon\right) \to 1$$

as $n \to \infty$, because if $\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)$ had positive measure in the limit, then the region $\Omega \setminus N(y)$ will contain data points from $\mathcal{X}(k)$ with positive probability for sufficiently large $k \geq n$ for each $y \in \lim_{n\to\infty} \Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)$. So y can not be in $\lim_{n\to\infty} \Gamma_1(\mathcal{X}(n), N)$, which is a contradiction. Hence from Proposition 5.6 of Karr (1992), the desired result follows. □

Remark 4.8 In Lemma 4.6 and Theorem 4.7, notice that we define the set $\mathcal{X}(n)$ recursively so that at each step a new point is added to the data set in the previous step. These results do not necessarily hold for the general set $\mathcal{X}_n$ (which is not defined recursively). The recursive nature of the data set $\mathcal{X}(n)$ is crucial for the results to hold. Furthermore, the result in Lemma 4.6 also holds for deterministic $\mathcal{X}(n)$ which are defined recursively as in the hypothesis of the lemma. However the result in Theorem 4.7 is an a.s. convergence result, so it holds for a random set $\mathcal{X}(n)$ as defined therein, but does not necessarily hold for deterministic $\mathcal{X}(n)$. A counterexample for the deterministic case can be constructed easily.

Note however that $\Gamma_1(\mathcal{X}_n, N)$ is neither strictly decreasing nor non-increasing provided that $\mathscr{R}_S(N) \neq \Omega$, because we might have $\Gamma_1(\mathcal{X}_m, N) \nsubseteq \Gamma_1(\mathcal{X}_n, N)$ for some $m > n$. Because it is possible to have a situation where $\cap_{i=1}^{n} \Gamma_1(X_i, N)$ does not contain $\cap_{j=1}^{m} \Gamma_1\left(X'_j, N\right)$, where $\mathcal{X}_m = \{X'_1, X'_2, \ldots, X'_m\}$ and $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ and $m > n$. Nevertheless, the following two results hold.

Proposition 4.9 *Suppose $\Omega \setminus \mathscr{R}_S(N)$ has positive measure. For positive integers m, n with $m > n$, let $\mathcal{X}_n$ and $\mathcal{X}_m$ be two samples from F on Ω. Then $\mu(\Gamma_1(\mathcal{X}_m, N)) \leq^{ST} \mu(\Gamma_1(\mathcal{X}_n, N))$ where $\leq^{ST}$ stands for "stochastically smaller than".*

Proof Recall that for $X \sim F$ and $Y \sim G$, we have $X \leq^{ST} Y$ if $F(x) \geq G(x)$ for all x with strict inequality holding for at least one x. Let $m > n$ and $\mathcal{X}_n$ and $\mathcal{X}_m$ be two samples from F. Then $\mu(\Gamma_1(\mathcal{X}_m, N))$ is more often smaller than $\mu(\Gamma_1(\mathcal{X}_n, N))$. Hence $P[\mu(\Gamma_1(\mathcal{X}_m, N)) \leq \mu(\Gamma_1(\mathcal{X}_n, N))] \geq 1/2$ which only shows stochastic precedence (Boland et al. 2004).

Now, let $t \in \big(\mu(\mathscr{R}_S(N)), \mu(\Omega)\big)$, then $\mu(\Gamma_1(\mathcal{X}_m, N)) \leq t$ happens more often than $\mu(\Gamma_1(\mathcal{X}_n, N)) \leq t$, hence $P(\mu(\Gamma_1(\mathcal{X}_m, N)) \leq t) \geq P(\mu(\Gamma_1(\mathcal{X}_n, N)) \leq t)$; that is, $F_m(t) \geq F_n(t)$, where $F_k(\cdot)$ is the distribution function for $\mu(\Gamma_1(\mathcal{X}_k, N))$ for $k = m, n$. For $t < \mu(\mathscr{R}_S(N))$ or $t > \mu(\Omega)$, we have $F_k(t) = 0$ for $k = m, n$. Letting $N_n := |\mathcal{X}_n \setminus \mathscr{R}_S(N)|$ and $N_m := |\mathcal{X}_m \setminus \mathscr{R}_S(N)|$, then $P(N_n \neq N_m) > 0$ since $\mu(\Omega \setminus \mathscr{R}_S(N)) > 0$. In fact, $P(N_m > N_n) \geq 1/2$. But if $F_m(t) = F_n(t)$ for all t were the case, then $P(N_n = N_m) = 1$ would have held, which is a contradiction. Hence we have $\mu(\Gamma_1(\mathcal{X}_m, N)) \leq^{ST} \mu(\Gamma_1(\mathcal{X}_n, N))$. □

Theorem 4.10 *Let $\{\mathcal{X}_n\}_{n=1}^{\infty}$ be a sequence of samples each of which is of size n from distribution F with support on Ω. Then $\Gamma_1(\mathcal{X}_n, N) \xrightarrow{p} \mathscr{R}_S(N)$ in the sense that $\mu(\Gamma_1(\mathcal{X}_n, N) \setminus \mathscr{R}_S(N)) \xrightarrow{p} 0$ as $n \to \infty$.*

Proof Suppose we have a sequence $\{\mathcal{X}_n\}_{n=1}^{\infty}$ as in the theorem. By Proposition 4.3, $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}_n, N)$ for each n. Let $\Lambda_n := \Gamma_1(\mathcal{X}_n, N) \setminus \mathscr{R}_S(N)$. If Λ_n has zero measure in the limit as $n \to \infty$, then the result follows trivially. Otherwise, if Λ_n had positive measure in the limit, for each $y \in \lim_{n\to\infty} \Lambda_n$, the region $\Omega \setminus N(y)$ would have positive measure, then $\mathcal{X}_n \cap [\Omega \setminus N(y)] \neq \emptyset$ with positive probability for sufficiently large n, then $y \notin \Gamma_1(\mathcal{X}_n, N)$, which is a contradiction. □

Theorem 4.11 *Let $\gamma_n(N)$ be the domination number of the PCD based on proximity map N and the data set $\mathcal{X}_n$ which is a random sample from a distribution F with support $\mathcal{S}(F) \subseteq \Omega$. If the superset region for N and $\mathcal{S}(F) \cap \mathscr{R}_S(N)$ have positive measure (that is, e.g., $\mu(\mathscr{R}_S(N)) > 0$), then $P(\gamma_n(N) = 1) \to 1$ as $n \to \infty$.*

Proof Suppose $\mu(\mathscr{R}_S(N)) > 0$ and $\mu(\mathcal{S}(F) \cap \mathscr{R}_S(N)) > 0$ and $\mathcal{X}_n$ is as in the theorem. Notice that if there is at least one data point in $\mathscr{R}_S(N)$ then $\gamma_n(N) = 1$, because any point $x \in \mathscr{R}_S(N)$ will have $N(x) = \Omega$, so $P(\mathcal{X}_n \cap \mathscr{R}_S(N) \neq \emptyset) \leq P(\gamma_n(N) = 1)$. Now, $P(\mathcal{X}_n \cap \mathscr{R}_S(N) \neq \emptyset) = 1 - P(\mathcal{X}_n \cap \mathscr{R}_S = \emptyset) = 1 - \left(\frac{P(X\in\Omega) - P(X\in\mathscr{R}_S(N))}{P(X\in\Omega)}\right)^n = 1 - \big(1 - P(X \in \mathscr{R}_S(N))\big)^n$, which goes to 1 as $n \to \infty$, since $0 < P(X \in \mathscr{R}_S(N)) \leq 1$. Therefore $P(\gamma_n(N) = 1) \to 1$ as $n \to \infty$. □

Remark 4.12 (Relative Arc Density) The *relative arc density* of a digraph $D = (\mathcal{V}, \mathcal{A})$ of order $|\mathcal{V}| = n$, denoted as $\rho(D)$, is defined as $\rho(D) := \frac{|\mathcal{A}|}{n(n-1)}$ where $|\cdot|$ denotes the cardinality of sets (Janson et al. 2000). Thus $\rho(D)$ represents the ratio of the number of arcs in the digraph D to the number of arcs in the complete symmetric digraph of order n, which is $n(n-1)$.

Theorem 4.13 *Let ρ_n be the relative arc density of the PCD based on the proximity map N and data set $\mathcal{X}_n$. If the support of the joint distribution of $\mathcal{X}_n$ is subset of the superset region for any type of proximity map, then we have $\rho_n = 1$ a.s.*

Proof Suppose the support of $\mathcal{X}_n$ is a subset of the superset region, then the corresponding digraph is complete with $n(n-1)$ arcs. Hence the relative arc density is 1 with probability 1. □

Ceyhan et al. (2006) demonstrates the asymptotic normality of the relative arc density of the proportional-edge PCD. The proximity region and Γ_1-region play a crucial role in finding the moments of this asymptotic normal distribution.

4.2 Superset Regions for Proportional-Edge Proximity Maps

Let $\mathscr{R}_S\left(N^r_{PE}, M_C\right)$ be the superset region for N^r_{PE} based on M_C-vertex regions constructed using the median lines. For $r < 3/2$, we have $\mathscr{R}_S(N^r_{PE}, M_C) = \emptyset$; for $r > 3/2$, the region $\mathscr{R}_S(N^r_{PE}, M_C)$ has positive area; and for $r = 3/2$, we have $\mathscr{R}_S(N^r_{PE}, M_C) = \{M_C\}$. Therefore, $r = 3/2$ is the threshold for the superset region to be nonempty. Furthermore, $r = 3/2$ is the only value at which the asymptotic distribution of the domination number of the PCD based on $N^r_{PE}(\cdot, M_C)$ is nondegenerate (Ceyhan and Priebe 2005).

Let $\mathscr{R}^{\perp}_S\left(N^r_{PE}, M\right)$ be the superset region for N^r_{PE} based on M-vertex regions with orthogonal projections. Then the superset region with the incenter $\mathscr{R}^{\perp}_S\left(N^2_{PE}, M_I\right)$ is as in Fig. 4 (left). Let M_i be the midpoint of edge e_i for $i = 1, 2, 3$. Then $T(M_1, M_2, M_3) \subseteq \mathscr{R}^{\perp}_S(N^2_{PE}, M_I)$ for all $T(\mathcal{Y}_3)$ with equality holding when $T(\mathcal{Y}_3)$ is an equilateral triangle.

For $N^2_{PE}(\cdot, M_C)$ which is constructed using the median lines, we have $\mathscr{R}_S\left(N^2_{PE}, M_C\right) = T(M_1, M_2, M_3)$, and for $N^2_{PE}(\cdot, M_C)$ which is constructed by using the orthogonal projections, we have $T(M_1, M_2, M_3) \subseteq \mathscr{R}^{\perp}_S\left(N^2_{PE}, M_C\right)$ with equality holding when $T(\mathcal{Y}_3)$ is an equilateral triangle.

In $T(\mathcal{Y}_3)$, consider the lines $\zeta_i(r, x)$ such that $d(\mathsf{y}_i, e_i) = r\, d(\zeta_i(r, x), \mathsf{y}_i)$ for $i \in \{1, 2, 3\}$. Note that if $r > 3/2$, then the straight lines $\zeta_i(r, x)$ yield a triangle similar to $T(M_1, M_2, M_3)$; if $r = 3/2$, then the straight lines $\zeta_i(r, x)$ intersect at the center of mass; and if $1 \le r < 3/2$, then the straight lines $\zeta_i(r, x)$ yield another

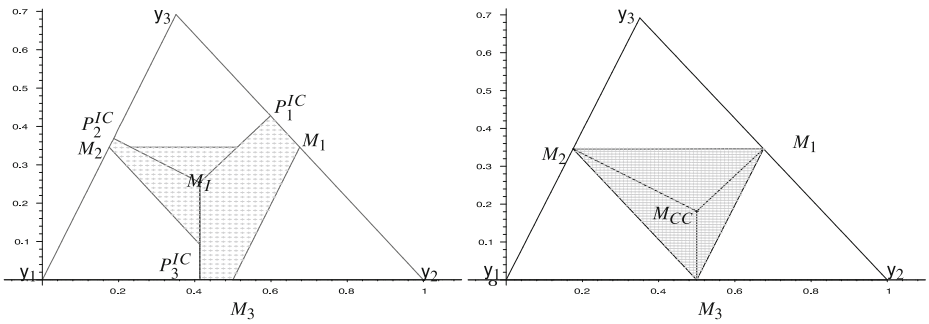

Fig. 4 Superset regions $\mathscr{R}^{\perp}_S\left(N^2_{PE}, M_I\right)$ (*left*) and $\mathscr{R}^{\perp}_S\left(N^2_{PE}, M_{CC}\right)$ (*right*) in the basic triangle T_b

triangle which is denoted as $\mathscr{T}^r$ and is similar to $T(\mathcal{Y}_3)$. The superset region $\mathscr{R}_S(N^r_{PE}, M)$ can be described as follows. $\mathscr{R}_S(N^r_{PE}, M) = \cup_{i=1}^3[\mathscr{R}_S(N^r_{PE}, M) \cap R_M(\mathsf{y}_i)]$ where $\mathscr{R}_S(N^r_{PE}, M) \cap R_M(\mathsf{y}_i) = \{z \in R_M(\mathsf{y}_i) : d(z, e_i) \le d(\zeta_i(r, x), e_i)\}$ for $i = 1, 2, 3$. See Fig. 4 with $M = M_{CC}$ and $M = M_I$ where the vertex regions are constructed using the orthogonal projections from M to the edges. For $r = 3/2$, the region $\mathscr{R}_S(N^r_{PE}, M)$ is guaranteed to be nonempty, since $M_C \in \mathscr{R}_S(N^r_{PE}, M)$. For $r > 3/2$, the region $\mathscr{R}_S(N^r_{PE}, M)$ is guaranteed to have positive measure. For $1 \le r < 3/2$, it follows that $\mathscr{R}_S(N^r_{PE}, M)$ may or may not be empty, depending on the location of M.

The functional form of the superset region, $\mathscr{R}_S\left(N^r_{PE}, M\right)$, is given by

$$\begin{aligned}\mathscr{R}_S\left(N^r_{PE}, M\right) = &\left\{(x, y) \in R_M(\mathsf{y}_1) : y \ge \frac{c_2\,(1 - r\,x)}{r\,(1 - c_1)}\right\}\\ &\bigcup \left\{(x, y) \in R_M(\mathsf{y}_2) : y \ge \frac{c_2\,(r\,(x - 1) + 1)}{r\,c_1}\right\}\\ &\bigcup \left\{(x, y) \in R_M(\mathsf{y}_3) : y \le c_2 \frac{r - 1}{r}\right\},\end{aligned}$$

and the functional form of $T(M_1, M_2, M_3)$ in T_b is given by

$$T(M_1, M_2, M_3) = \left\{(x, y) \in T_b : y \le \frac{c_2}{2};\ y \ge \frac{c_2\,(-1 + 2\,x)}{2\,c_1};\ y \ge \frac{c_2\,(1 - 2\,x)}{2\,(1 - c_1)}\right\}.$$

Recall that for $1 \le r < 3/2$, the triangle $\mathscr{T}^r$ is bounded by the lines $\zeta_i(r, x)$ for $i = 1, 2, 3$. See Fig. 5 for $\mathscr{T}^r$ with $r = \sqrt{2}$. The functional form of $\mathscr{T}^r$ in T_b is

$$\begin{aligned}\mathscr{T}^r &= \left\{(x, y) \in T_b : y \ge \frac{c_2\,(r - 1)}{r};\ y \le \frac{c_2\,(1 - r\,x)}{r\,(1 - c_1)};\ y \le \frac{c_2\,(r\,(x - 1) + 1)}{r\,c_1}\right\}\\ &= T\left(\left(\frac{(r - 1)\,(1 + c_1)}{r}, \frac{c_2\,(r - 1)}{r}\right), \left(\frac{2 - r + c_1\,(r - 1)}{r}, \frac{c_2\,(r - 1)}{r}\right),\right.\\ &\qquad \left.\left(\frac{c_1\,(2 - r) + r - 1}{r}, \frac{c_2\,(r - 2)}{r}\right)\right). \qquad (3)\end{aligned}$$

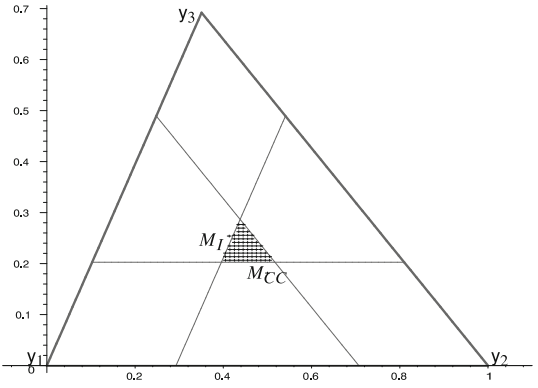

Fig. 5 An illustration of the triangle $\mathscr{T}^r$ with $r = \sqrt{2}$ in the basic triangle T_b

There is a crucial difference between $\mathscr{T}^r$ and $T(M_1, M_2, M_3)$: $T(M_1, M_2, M_3) \subseteq \mathscr{R}_S\left(N^r_{PE}, M\right)$ for all M and $r \geq 2$, but $(\mathscr{T}^r)^o$ and $\mathscr{R}_S\left(N^r_{PE}, M\right)$ are mutually exclusive for all M and r.

If $M \in (\mathscr{T}^r)^o$, then $\mathscr{R}_S\left(N^r_{PE}, M\right) = \emptyset$; if $M \in \partial(\mathscr{T}^r)$, then $\mathscr{R}_S\left(N^r_{PE}, M\right) = \{M\}$; and if $M \notin \mathscr{T}^r$, then $\mathscr{R}_S\left(N^r_{PE}, M\right)$ has positive area. See Fig. 4 (right) for an example of $\mathscr{R}^{\perp}_S(N^r_{PE}, M)$ with $r = 2$ and $M = M_{CC}$ (i.e., circumcenter). The triangle $\mathscr{T}^r$ defined above plays a crucial role in the analysis of the distribution of the domination number of the PCD based on proportional-edge proximity maps. The superset region $\mathscr{R}_S\left(N^r_{PE}, M\right)$ is important for both the domination number and the relative density of the corresponding PCDs.

The domination number of the PCD based on $N^r_{PE}(\cdot, M)$ is geometry invariant, if M-vertex regions are constructed with $M \in T(\mathcal{Y}_3)^o$ by using the extensions of the line segments joining y to M for all $\mathsf{y} \in \mathcal{Y}_3$. But for example when the vertex regions are constructed by orthogonal projections, the domination number of the PCD associated with $N^r_{PE}(\cdot, M)$ is not geometry invariant (Ceyhan 2009a), hence such vertex regions are not considered henceforth.

In $\mathbb{R}^d$ with $d > 2$, the superset region, provided that it is non-empty, lies in the simplex $\mathfrak{S}(\mathcal{Y}_{d+1})$. In $\mathfrak{S}(\mathcal{Y}_{d+1})$, consider the hyperplanes $\widetilde{\zeta}_i(r, x)$ such that $d(\mathsf{y}_i, \varphi_i) = r\, d\left(\widetilde{\zeta}_i(r, x), \mathsf{y}_i\right)$ for $i \in \{1, 2, 3\}$. The superset region $\mathscr{R}_S(N^r_{PE}, M)$ can be constructed as follows. $\mathscr{R}_S(N^r_{PE}, M) = \cup_{i=1}^{d+1}\left[\mathscr{R}_S(N^r_{PE}, M) \cap R_M(\mathsf{y}_i)\right]$ where $\mathscr{R}_S(N^r_{PE}, M) \cap R_M(\mathsf{y}_i) = \left\{z \in R_M(\mathsf{y}_i) : d(z, \varphi_i) \leq d\left(\widetilde{\zeta}_i(r, x), \varphi_i\right)\right\}$ for $i = 1, 2, \ldots, (d+1)$. We conjecture that if $r > (d+1)/d$, then the hyperplanes $\widetilde{\zeta}_i(r, x)$ yield a simplex which is similar to the simplex with vertices $(M_1, M_2, \ldots, M_{d+1})$ and is a subset of the superset region; if $r = (d+1)/d$, then the hyperplanes $\widetilde{\zeta}_i(r, x)$ intersect at the center of mass; and if $r < (d+1)/d$, then the hyperplanes $\widetilde{\zeta}_i(r, x)$ yield another simplex which is similar to $\mathfrak{S}(\mathcal{Y}_{d+1})$ and is the counterpart of $\mathscr{T}^r$ in $\mathbb{R}^2$.

4.3 Γ_1-Regions for Proportional-Edge Proximity Maps

Since $\Gamma_1(\mathcal{X}_n, N) = \bigcap_{i=1}^{n} \Gamma_1(x_i, N)$ for a given realization of the data set $\mathcal{X}_n$, first we describe the region $\Gamma_1(x, N)$ for $x \in \mathcal{X}_n$ and then describe the region $\Gamma_1(\mathcal{X}_n, N)$.

For $N^r_{PE}(\cdot, M)$, the Γ_1-region denoted as $\Gamma^r_1(\cdot, M)$ is constructed as follows; see also Fig. 6. Let $\xi_i(r, x)$ be the line parallel to e_i such that $\xi_i(r, x) \cap T(\mathcal{Y}_3) \neq \emptyset$ and $r\, d(\mathsf{y}_i, \xi_i(r, x)) = d(\mathsf{y}_i, \ell(\mathsf{y}_i, x))$ for $i \in \{1, 2, 3\}$. Then

$$\Gamma^r_1(x, M) = \bigcup_{i=1}^{3}\left[\Gamma^r_1(x, M) \cap R_M(\mathsf{y}_i)\right]$$

where

$$\Gamma^r_1(x, M) \cap R_M(\mathsf{y}_i) = \{z \in R_M(\mathsf{y}_i) : d(\mathsf{y}_i, \ell(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \xi_i(r, x)\} \text{ for } i \in \{1, 2, 3\}.$$

Notice that $r \geq 1$ implies $x \in \Gamma^r_1(x, M)$ for each x. Furthermore, $\lim_{r\to\infty} \Gamma^r_1(x, M) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$ and so we define $\Gamma^{r=\infty}_1(x, M) = T(\mathcal{Y}_3)$ for all such x. For $x \in \mathcal{Y}_3$, $\Gamma^r_1(x, M) = \{x\}$ for all $r \in [1, \infty]$.

The functional form of $\Gamma^r_1(x = (x_0, y_0), M)$ in the basic triangle T_b is given by

$$\Gamma^r_1(x = (x_0, y_0), M) = \bigcup_{i=1}^{3}\left[\Gamma^r_1(x = (x_0, y_0), M) \cap R_M(y_i)\right]$$

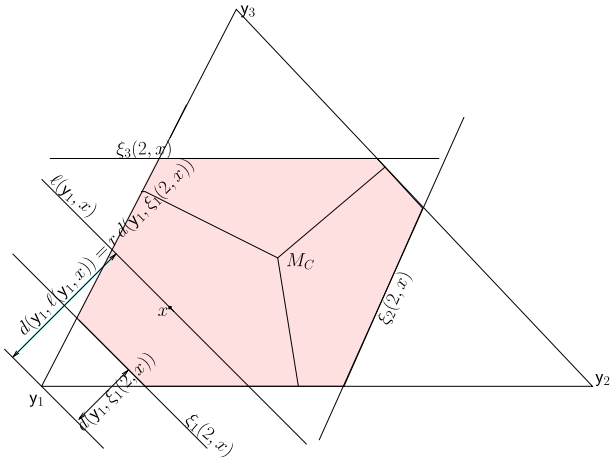

Fig. 6 Construction of the Γ_1-region (*shaded region*), $\Gamma_1^r(x, M_C)$ with $r = 2$ for an $x \in R_{M_C}(\mathsf{y}_1)$

where

$$\Gamma_1^r(x = (x_0, y_0), M) \cap R_M(\mathsf{y}_1) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \geq \frac{y_0}{r} - \frac{c_2\,(r\,x - x_0)}{(1 - c_1)\,r}\right\},$$

$$\Gamma_1^r(x = (x_0, y_0), M) \cap R_M(\mathsf{y}_2) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \geq \frac{y_0}{r} - \frac{c_2\,(r\,(x - 1) + 1 - x_0}{c_1\,r}\right\},$$

$$\Gamma_1^r(x = (x_0, y_0), M) \cap R_M(\mathsf{y}_3) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \leq \frac{y_0 - c_2\,(1 - r)}{r}\right\}.$$

Notice that $\Gamma_1^r(x, M_C)$ is a convex hexagon for all $r \geq 2$ and $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$ (since for such an x, the region $\Gamma_1^r(x, M_C)$ is bounded by $\xi_i(r, x)$ and e_i for all $i \in \{1, 2, 3\}$, see also Fig. 6); else, it is either a convex hexagon or a non-convex polygon depending on the location of x and the value of r. The extension the associated Γ_1-region $\Gamma_1^r(\cdot, M)$ to higher dimensions is presented below.

Remark 4.14 We only present the extension of the Γ_1-region $\Gamma_1^r(\cdot, M = M_C)$ to higher dimensions. The extension for other M is similar. In $\mathbb{R}^d$ with $d > 2$, recall the simplex, $\mathfrak{S}(\mathcal{Y}_m)$, based on $d + 1$ points that do not lie on the same hyperplane. Furthermore, let $\varrho_i(r, x)$ be the hyperplane such that $\varrho_i(x) \cap \mathfrak{S}(\mathcal{Y}_m) \neq \emptyset$ and $r\,d(\mathsf{y}_i, \varrho_i(r, x)) = d(\mathsf{y}_i, \Upsilon(\mathsf{y}_i, x))$ for $i \in \{1, 2, \ldots, d + 1\}$. Then

$$\Gamma_1^r(x, M_C) \cap R_{M_C}(\mathsf{y}_i) = \{z \in R_{M_C}(\mathsf{y}_i) : d(\mathsf{y}_i, \Upsilon(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \varrho_i(r, x)\} \text{ for } i \in \{1, 2, 3\}.$$

Hence $\Gamma_1^r(x, M_C) = \bigcup_{i=1}^{d+1} [\Gamma_1^r(x, M_C) \cap R_{M_C}(\mathsf{y}_i)]$. Furthermore, it is easy to see that

$$\Gamma_1^r(\mathcal{X}_n, M_C) = \bigcap_{i=1}^{d+1} \Gamma_1^r\left(X_{\varphi_i}(n), M_C\right),$$

where $X_{\varphi_i}(n)$ is one of the closest points in $\mathcal{X}_n \cap R_{M_C}(\mathsf{y}_i)$ to face φ_i.

So far, we have described the Γ_1-region for a point in $x \in T(\mathcal{Y}_3)$. For a set $\mathcal{X}_n$ of size n in $T(\mathcal{Y}_3)$, the region $\Gamma_1^r(\mathcal{X}_n, M)$ can be determined by the edge extrema.

Definition 4.15 (Edge Extrema) The (closest) *edge extrema* of a set B in $T(\mathcal{Y}_3)$ are the points closest to the edges of $T(\mathcal{Y}_3)$, denoted $x_{[i]}$ for $i \in \{1, 2, 3\}$; that is, $x_{[i]} \in \operatorname{arginf}_{x \in B} d(x, e_i)$.

Note that if $B = \mathcal{X}_n$ is a set of Ω-valued random variables of size n from F, then the edge extrema, denoted $X_{[i]}$, are also random variables.

Theorem 4.16 *Let B be any set of n distinct points in $T(\mathcal{Y}_3)$ and $x_{[i]} \in \operatorname{arginf}_{x \in B} d(x, e_i)$. For proportional-edge proximity maps with M-vertex regions, we have $\Gamma_1^r(B, M) = \bigcap_{i=1}^3 \Gamma_1^r\left(x_{[i]}, M\right)$.*

Proof Suppose $B = \{x_1, x_2, \ldots, x_n\}$ in $T(\mathcal{Y}_3)$. Note that

$$\Gamma_1^r(B, M) \cap R_M(\mathsf{y}_i) = \left[\bigcap_{i=1}^n \Gamma_1^r(x_i, M)\right] \bigcap R_M(\mathsf{y}_i),$$

and if $d(\mathsf{y}_i, \ell(\mathsf{y}_i, x)) \le d(\mathsf{y}_i, \ell(\mathsf{y}_i, x'))$ then $N_{PE}^r(x, M) \subseteq N_{PE}^r(x', M)$ for all $x, x' \in R_M(\mathsf{y}_i)$. Further, by definition $x_{[i]} \in \operatorname{argmax}_{x \in B} d(\mathsf{y}_i, \xi_i(r, x))$, so

$$\Gamma_1^r(B, M) \cap R_M(\mathsf{y}_i) = \Gamma_1^r\left(x_{[i]}, M\right) \cap R_M(\mathsf{y}_i) \text{ for } i \in \{1, 2, 3\}.$$

Furthermore, $\Gamma_1^r(B, M) = \bigcup_{i=1}^3 \left[\Gamma_1^r\left(x_{[i]}, M\right) \cap R_M(\mathsf{y}_i)\right]$, and

$$\Gamma_1^r\left(x_{[i]}, M\right) \cap R_M(\mathsf{y}_i) = \left[\bigcap_{j=1}^3 \Gamma_1^r\left(x_{[j]}, M\right)\right] \bigcap R_M(\mathsf{y}_i) \text{ for } i \in \{1, 2, 3\}.$$

Combining these two results, we obtain $\Gamma_1^r(B, M) = \bigcap_{j=1}^3 \Gamma_1^r\left(x_{[j]}, M\right)$. □

Then $\Gamma_1^r(\mathcal{X}_n, M) = \bigcap_{i=1}^3 \Gamma_1^r\left(X_{[i]}, M\right)$, where $X_{[i]}$ is the closest point to edge e_i which is the edge opposite vertex y_i, for $i = 1, 2, 3$. So $\Gamma_1^r(\mathcal{X}_n, M) \cap R_M(\mathsf{y}_i) = \{z \in R_M(\mathsf{y}_i) : d(\mathsf{y}_i, \ell(\mathsf{y}_i, z)) \ge d(\mathsf{y}_i, \xi_i(r, x_{[i]}))\}$ for $i = 1, 2, 3$.

For $r \ge 3/2$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, we have $\Gamma_1^r(\mathcal{X}_n, M) \ne \emptyset$ a.s., since $\mathscr{R}_S\left(N_{PE}^r, M\right) \ne \emptyset$ and $\mathscr{R}_S\left(N_{PE}^r, M\right) \subseteq \Gamma_1^r(\mathcal{X}_n, M)$.

Now, for $n > 1$, let $X_{[i]} = x_{[i]} = (u_i, w_i)$ be given for $i \in \{1, 2, 3\}$, be the edge extrema in a given realization of $\mathcal{X}_n$. The functional form of Γ_1-region in T_b is given by

$$\Gamma_1^r(\mathcal{X}_n, M) = \bigcup_{i=1}^3 \left[\Gamma_1^r(\mathcal{X}_n, M) \cap R_M(\mathsf{y}_i)\right]$$

where

$$\Gamma_1^r(\mathcal{X}_n, M) \cap R_M(\mathsf{y}_1) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \ge \frac{w_1}{r} - \frac{c_2(r\,x - u_1)}{(1 - c_1)\,r}\right\},$$

$$\Gamma_1^r(\mathcal{X}_n, M) \cap R_M(\mathsf{y}_2) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \ge \frac{w_2}{r} - \frac{c_2(r\,(x - 1) + 1 - u_2}{c_1\,r}\right\},$$

$$\Gamma_1^r(\mathcal{X}_n, M) \cap R_M(\mathsf{y}_3) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \le \frac{w_3 - c_2(1 - r)}{r}\right\}.$$

However, for $1 \le r < 3/2$, the region $\Gamma_1^r(x, M) \cap R_M(\mathsf{y}_i)$ might be empty for some $i \in \{1, 2, 3\}$. Furthermore, if $M \in (\mathscr{T}^r)^o$ with $1 \le r < 3/2$, then $\Gamma_1^r(\mathcal{X}_n, M)$ will be empty with probability 1 as $n \to \infty$. In such a case, there is no Γ_1-region to construct. To determine whether the Γ_1-region is empty or not, it suffices to check the intersection of the Γ_1-regions of the edge extrema. If $M \notin (\mathscr{T}^r)^o$, then the Γ_1-region is guaranteed to be nonempty.

Remark 4.17

- For $r_1 < r_2$, $\Gamma_1^{r_1}(x, M) \subseteq \Gamma_1^{r_2}(x, M)$ for all $x \in T(\mathcal{Y}_3)$ with equality holding only when $x \in \mathcal{Y}_3$; and $\Gamma_1^{r_1}(\mathcal{X}_n, M) \subseteq \Gamma_1^{r_2}(\mathcal{X}_n, M)$ with equality holding only when $\mathcal{X}_n \subseteq \mathcal{Y}_3$ or $\Gamma_1^r(\mathcal{X}_n, M) = \emptyset$ for $r \in \{r_1, r_2\}$.
- Suppose X and X' are iid from a continuous distribution F whose support is $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. Then for $r_1 < r_2$, we have $A\left(\Gamma_1^{r_1}(X, M)\right) \preceq^{ST} A\left(\Gamma_1^{r_2}(X', M)\right)$.
- Suppose $\mathcal{X}_n$ and $\mathcal{X}_n'$ are two random samples from a continuous distribution F whose support is $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. Then for $r_1 < r_2$, we have $A(\Gamma_1^{r_1}(\mathcal{X}_n, M)) \preceq^{ST} A(\Gamma_1^{r_2}(\mathcal{X}_n', M))$.

In Section 4.1, we have investigated the behavior of $\Gamma_1(\mathcal{X}_n, N)$ for general proximity maps in Ω. The assertions for the specific proximity maps we consider will be stronger. For example, we have a stronger result than the one in Proposition 4.3 in the sense that, $\mathscr{R}_S(N)$ is a proper subset of $\Gamma_1(\mathcal{X}_n, N)$ (i.e., $\mathscr{R}_S(N) \subsetneq \Gamma_1(\mathcal{X}_n, N)$) as shown below.

Proposition 4.18 *For each type of proximity map $N \in \left\{N_S, N_{AS}, N_{PE}^r\right\}$ and any random sample $\mathcal{X}_n = \left\{X_1, X_2, \ldots, X_n\right\}$ from a continuous distribution F on $T(\mathcal{Y}_3)$, if $\mathscr{R}_S(N) \neq \emptyset$, then $\mathscr{R}_S(N) \subsetneq \Gamma_1(\mathcal{X}_n, N)$ a.s. for each $n < \infty$.*

Proof We have shown that $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}_n, N)$ (see Proposition 4.3). Moreover, $\mathscr{R}_S(N) \neq \emptyset$ for N_{AS} and N_{PE}^r with $r > 3/2$. For these proximity regions, $\mathscr{R}_S(N) = \Gamma_1(\mathcal{X}_n, N)$ with probability 0 for each finite n since

(i) for $N \in \{N_S, N_{AS}\}$, we have $\Gamma_1(\mathcal{X}_n, N) = \mathscr{R}_S(N)$ iff $X_{\mathsf{y}}(n) = \mathsf{y}$ for each $\mathsf{y} \in \mathcal{Y}_3$ which happens with probability 0,
(ii) for N_{PE}^r with $r > 3/2$, we have $\Gamma_1^r(\mathcal{X}_n, M) = \mathscr{R}_S(N, M)$ iff $X_{[i]} \in e_i$ for each $i \in \{1, 2, 3\}$ which happens with probability 0.

Furthermore, for N_{PE}^r with $1 \le r < 3/2$, we have $\mathscr{R}_S\left(N_{PE}^r, M\right) \neq \emptyset$ iff $M \notin (\mathscr{T}^r)^o$, say $M = (m_x, m_y)$ is such that $d(\zeta_2(m_x, x), \mathsf{y}_2) < d(\mathsf{y}_2, e_2)/r$. Then $\Gamma_1^r(\mathcal{X}_n, M) = \mathscr{R}_S\left(N_{PE}^r, M\right)$ iff $X_{[2]} \in e_2$ which happens with probability 0. Similarly the same result also holds for edges e_1 and e_3. □

Note that, if $\mathscr{R}_S(N) = \emptyset$ and $\mathcal{X}_n$ is a random sample from a continuous distribution on $T(\mathcal{Y}_3)$, then $\Gamma_1(\mathcal{X}_n, N) = \emptyset$ a.s. in the limit as $n \to \infty$. In particular, this holds for $N_{PE}^r(\cdot, M)$ with $1 \le r < 3/2$ and $M \in (\mathscr{T}^r)^o$. Lemma 4.6 holds as stated. In Lemma 4.6, we have shown that $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing. Furthermore, for the proximity regions $\left\{N_S, N_{AS}, N_{PE}^r\right\}$, we can state that $\Gamma_1(\mathcal{X}(n+1), N) \subsetneq \Gamma_1(\mathcal{X}(n), N)$ with positive probability, since the new point in $\mathcal{X}(n+1)$ has positive probability to fall closer to the subset of $T(\mathcal{Y}_3)$ that defines $\mathscr{R}_S(N)$ (e.g., $\partial(T(\mathcal{Y}_3))$). The general

results in Theorems 4.7 and 4.10 and Proposition 4.9 hold for the proximity maps $\{N_S, N_{AS}, N^r_{PE}\}$ also.

4.4 Expected Value of the Measure of Γ_1-Regions

Let $\lambda(\cdot)$ be the Lebesgue measure on $\mathbb{R}^d$ with $d \geq 1$. In $\mathbb{R}$ with $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\}$, let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$. We can assume that $(\mathsf{y}_1, \mathsf{y}_2) = (0, 1)$ due to the geometry invariance for uniform data in $\mathbb{R}$. Then, it follows that $\Gamma_1(\mathcal{X}_n, N_S) = [X_{n:n}/2, (1 + X_{1:n})/2]$. So $\lambda(\Gamma_1(\mathcal{X}_n, N_S)) = |\Gamma_1(\mathcal{X}_n, N_S)| = (1 + X_{1:n} - X_{n:n})/2$. Hence the expected length of the Γ_1-region is

$$\begin{aligned}\mathbf{E}[\lambda(\Gamma_1(\mathcal{X}_n, N_S)] &= \mathbf{E}\left[\frac{1 + X_{1:n} - X_{n:n}}{2}\right] \\ &= \frac{1 + \mathbf{E}[X_{1:n}] - \mathbf{E}[X_{n:n}]}{2} = \frac{1 + \frac{1}{n+1} - \frac{n}{n+1}}{2} \\ &= \frac{1}{n+1} \to 0\end{aligned}$$

as $n \to \infty$, since $\mathbf{E}[X_{k:n}] = \frac{k}{n+1}$ for iid uniform X_i in $(0, 1)$.

4.4.1 The Limit of Expected Area of $\Gamma^r_1(\mathcal{X}_n, M)$

In $\mathbb{R}^2$, with three non-collinear points $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$, let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. Then for $r > 3/2$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, we have $A(\Gamma^r_1(\mathcal{X}_n, M)) > 0$ a.s. for all $n < \infty$. The region $\Gamma^r_1(\mathcal{X}_n, M)$ is determined by the (closest) edge extrema $X_{[i]} \in \operatorname{argmin}_{X \in \mathcal{X}_n} d(X, e_i)$ for $i \in \{1, 2, 3\}$. So, to find the expected area of $\Gamma^r_1(\mathcal{X}_n, M)$, we need to find the expected locus of $X_{[i]}$; i.e., the expected distance of $X_{[i]}$ from e_i. For example, for $\mathcal{X}_n$ a random sample from a continuous distribution F, $\operatorname{argmin}_{X \in \mathcal{X}_n} d(X, e_i)$ is unique a.s., and if $d(X_{[i]}, e_i) = u$, then $X_{[i]}$ falls on a line parallel to e_i whose distance from e_i is u a.s. for $i = 1, 2, 3$.

Lemma 4.19 *Let $D_i(n) := d(X_{[i]}, e_i)$ for $i \in \{1, 2, 3\}$ and $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. Then $\mathbf{E}[D_i(n)] \to 0$ (i.e., the expected locus of $X_{[i]}$ converges to edge e_i) as $n \to \infty$ for each $i \in \{1, 2, 3\}$.*

Proof Suppose $Z_i = (X_i, Y_i) \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$. Then for $e = e_3$, it follows that $D_3(n) = Y_{1:n}$ which is the minimum y-coordinate of $Z_i \in \mathcal{X}_n$. First observe that $P(Y_i \leq y) = \frac{y(2c_2 - y)}{c_2^2}$, hence

$$F_Y(y) = \frac{y(2c_2 - y)}{c_2^2}\mathbf{I}(0 \leq y < c_2) + \mathbf{I}(y \geq c_2).$$

So the pdf of Y_i is

$$f_Y(y) = 2\frac{c_2 - y}{c_2^2}\mathbf{I}(0 \leq y \leq c_2).$$

Then the pdf of $Y_{1:n}$ is

$$f_{1:n}(y) = 2n(c_2 - y)\left(1 - \frac{y\,(2\,c_2 - y)}{c_2^2}\right)^{n-1} {c_2}^{-2}\mathbf{I}(0 \le y \le c_2).$$

Therefore,

$$\mathbf{E}\,[Y_{1:n}] = \int_0^{c_2} 2\,y\,n(c_2 - y)\left(1 - \frac{y\,(2\,c_2 - y)}{c_2^2}\right)^{n-1} {c_2}^{-2} dy = \frac{c_2}{2\,n+1} \to 0, \text{ as } n \to \infty.$$

Hence $\mathbf{E}\,[Y_{1:n}] = \mathbf{E}\,[D_3(n)] \to 0$. Similarly, as $n \to \infty$ $\mathbf{E}\,[D_i(n)] \to 0$ for $i \in \{1, 2\}$.

□

Theorem 4.20 *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$ and $M \in T(\mathcal{Y}_3)^o$. Then $\mathbf{E}\,[A(\Gamma_1^r(\mathcal{X}_n, M))] \to A(\mathscr{R}_S(N_{PE}^r, M))$ as $n \to \infty$.*

Proof Recall that for N_{PE}^r, we have $\Gamma_1^r(\mathcal{X}_n, M) = \bigcap_{i=1}^3 \Gamma_1^r\left(X_{[i]}, M\right)$. Moreover, $\Gamma_1^r(\mathcal{X}_n, M) = \mathscr{R}_S(N_{PE}^r, M)$ iff $X_{[i]} \in e_i$ for $i \in \{1, 2, 3\}$. In Lemma 4.19, we have shown that expected locus of $X_{[i]}$ converges to edge e as $n \to \infty$. Hence the expected locus of $\partial(\Gamma_1^r(\mathcal{X}_n, M)) \cap R_M(e_i)$ converges to the $\partial(\mathscr{R}_S(N_{PE}^r, M)) \cap R_M(e_i)$ for each $i \in \{1, 2, 3\}$. Hence

$$\mathbf{E}\,[A(\Gamma_1^r(\mathcal{X}_n, M)] \to A(\mathscr{R}_S(N_{PE}^r, M)) \text{ as } n \to \infty.$$

□

Remark 4.21 In particular,

i- $\mathbf{E}\left[A\left(\Gamma_1^{r=2}(\mathcal{X}_n, M_C)\right)\right] \to 1/4$ as $n \to \infty$, since $\mathscr{R}_S\left(N_{PE}^2, M_C\right) = T(M_1, M_2, M_3)$; and we have $\mathbf{E}\,[A(\Gamma_1^r(\mathcal{X}_n, M))] \to 0$ as $n \to \infty$ if $M \in \mathscr{T}^r$, since $\mathscr{R}_S\left(N_{PE}^r, M\right) = \emptyset$ for $M \in \mathscr{T}^r$. Furthermore, $\mathbf{E}\left[A\left(\Gamma_1^{r=3/2}(\mathcal{X}_n, M_C)\right)\right] \to 0$ since $\mathscr{R}_S\left(N_{PE}^{3/2}, M_C\right) = \{M_C\}$.

ii- We also have $\mathbf{E}\left[A\left(\Gamma_1^r(\mathcal{X}_n, M_C)\right)\right] \to 0$ for $r \in [1, 3/2)$ as $n \to \infty$. By careful geometric calculations, we get $\mathbf{E}\left[A\left(\Gamma_1^r(\mathcal{X}_n, M_C)\right)\right] \to \sqrt{3}\left[1 - 3/(2\,r)\right]^2$ for $r \in (3/2, 2]$. Furthermore, $\mathbf{E}\left[A\left(\Gamma_1^r(\mathcal{X}_n, M_C)\right)\right] \to \sqrt{3}/\left[4\,(1 - 3/r^2)\right]$ for $r \in (2, \infty]$, as $n \to \infty$.

Definition 4.22 Suppose $\mathcal{X}_n$ is a set of iid random variables from F with support $\mathcal{S}(F) \subseteq \Omega$. If over a sequence $\Omega_n \subseteq \Omega$ with $n = 1, 2, 3, \ldots$, X restricted to Ω_n, denoted $X|_{\Omega_n}$, has distribution F_n with $F_n(x) = F(x)/P_F(X \in \Omega_n)$ and $P_F(X \in \Omega_n) \to 1$ as $n \to \infty$, then we call F_n the *asymptotically accurate distribution* of X and Ω_n the *asymptotically accurate support* of F. If F has density f, then $f_n = f(x)/P_F(X \in \Omega_n)$ is called the *asymptotically accurate pdf* of X. In both cases, if we are concerned with asymptotic results, then for simplicity we will, respectively, use F and f for the asymptotically accurate distribution and pdf. Conditioning will be implied by stating that $X \in \Omega_n$ with probability 1, "as $n \to \infty$" or "for sufficiently large n".

We also derive the rate of convergence of $\mathbf{E}\,[A(\Gamma_1^r(\mathcal{X}_n, M_C))]$ for $r = 3/2$. First we provide a result that will be used in our derivation.

Theorem 4.23 *Let $\mathcal{X}_n$ be a random sample from F such that $B(\mathsf{y}_i, \varepsilon) \subseteq \mathcal{S}(F)$ for some $\varepsilon > 0$ and for each $i = 1, 2, 3$, and let $E_3(n)$ be the event that (closest) edge extrema are distinct. Then $P(E_3(n)) \to 1$ as $n \to \infty$.*

Proof First, we define two more events related to edge extrema as in the hypothesis of the theorem. Let $E_2(n)$ be the event that there are two distinct (closest) edge extrema and $E_1(n)$ be the event that all three edge extrema coincide. Clearly $P(E_1(n)) + P(E_2(n)) + P(E_3(n)) = 1$. It is trivial to see that $P(E_1(n)) \to 0$ as $n \to \infty$. Next we calculate the limit of $P(E_2(n))$ as $n \to \infty$.

Using the transformation $\phi_e : (x, y) \to (u, v)$ of Section 3.2, F becomes F_e with support $\mathcal{S}(F_e) \subseteq T_e$. Then $\phi_e(\mathcal{X}_n)$ becomes a random sample from F_e such that $B(\phi_e(\mathsf{y}_i), \varepsilon_e) \subseteq \mathcal{S}(F_e)$ for some $\varepsilon_e > 0$ and for each $i = 1, 2, 3$. First consider $P(X_{[2]} = X_{[3]})$. Given $X_{[2]} = X_{[3]} = (x, y)$ the remaining $n - 1$ points will lie in the shaded region in Fig. 7. The event $X_{[2]} = X_{[3]} = (X, Y)$ is equivalent to the event that

$$\mathcal{X}_n \subset S_R(X, Y) = \{(U, W) \in T_e : \ \ell(\mathsf{y}_1, (U, W)) \le \ell(\mathsf{y}_1, (X, Y)), \\ \ell(\mathsf{y}_3, (U, W)) \le \ell(\mathsf{y}_3, (X, Y))\}.$$

The pdf of such (X, Y) is $g(x, y) = n\, G(x, y)^{n-1}\, f(x, y)$ where $G(u, v) = P_F(X \in S_R(u, v))$. Note that $P(X_{[2]} = X_{[3]} = \mathsf{y}_1) = 0$ for all n, and $X_{[2]} = X_{[3]} \neq \mathsf{y}_1$ is equivalent to $d((X, Y), \mathsf{y}_1) > 0$. Let $\varepsilon > 0$, by Markov's inequality, we have $P(d((X, Y), \mathsf{y}_1) > \varepsilon) \le \mathbf{E}\left[d((X, Y), \mathsf{y}_1)\right]/\varepsilon = \mathbf{E}\left[\sqrt{X^2 + Y^2}\right]/\varepsilon$. Switching to the polar coordinates as $X = R\cos\theta$ and $Y = R\sin\theta$, we get $\sqrt{X^2 + Y^2} = R$. But,

$$\mathbf{E}[R] = \int_0^{\varepsilon} \int_0^{\pi/3} n\, r\, G(r, \theta)^{n-1}\, f(r, \theta) r\, dr\, d\theta.$$

The integrand is critical at $r = 0$, since for $r > 0$ it converges to zero as $n \to \infty$. So we use the Taylor series expansion around $r = 0$ as

$$f(r, \theta) = f(0, \theta) + \frac{\partial f(0, \theta)}{\partial r} r + O\left(r^2\right),$$
$$G(r, \theta) = G(0, \theta) + \frac{\partial G(0, \theta)}{\partial r} r + O\left(r^2\right) = 1 + \frac{\partial G(0, \theta)}{\partial r} r + O\left(r^2\right).$$

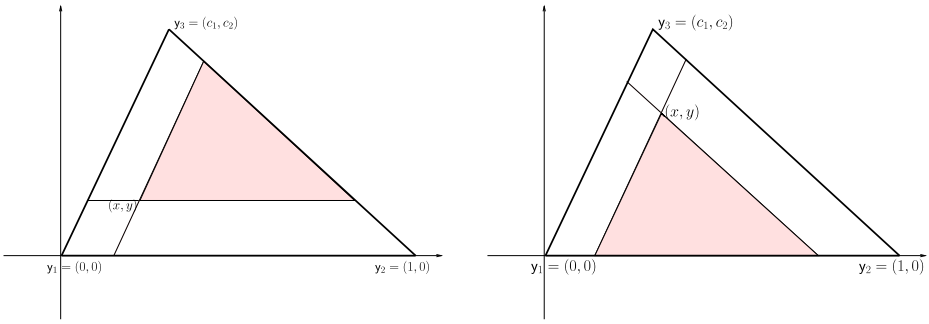

Fig. 7 The figure for $X_{[2]} = X_{[3]} = (x, y)$ (*left*) and $X_{[1]} = X_{[2]} = (x, y)$ (*right*)

Note that $\frac{\partial G(0,\theta)}{\partial r} < 0$, since area of $S_R(u, v)$ decreases as r increases for fixed θ. So let $r = w/n$, then

$$\begin{aligned}\mathbf{E}[R] &\approx \int_0^{n\varepsilon}\int_0^{\pi/3} n\frac{w}{n}\left(1+\frac{\partial G(0,\theta)}{\partial r}\frac{w}{n}+O\left(n^{-2}\right)\right)^{n-1}\\ &\quad\times\left(f(0,\theta)+\frac{\partial f(0,\theta)}{\partial r}\frac{w}{n}+O\left(n^{-2}\right)\right)\frac{w}{n^2}\,dw\,d\theta\\ &= \frac{1}{n^2}\int_0^{\infty}\int_0^{\pi/3} w^2\exp\left(\frac{\partial G(0,\theta)}{\partial r}w\right) f(0,\theta)w\,dw\,d\theta = O\left(n^{-2}\right).\end{aligned}$$

Hence $P(X_{[2]} = X_{[3]} \neq \mathsf{y}_1) \to 0$ as $n \to \infty$. Since $P(X_{[2]} = X_{[3]}) = P(X_{[2]} = X_{[3]} = \mathsf{y}_1) + P(X_{[2]} = X_{[3]} \neq \mathsf{y}_1)$ and $P(X_{[2]} = X_{[3]} = \mathsf{y}_1) = 0$, we have $P(X_{[2]} = X_{[3]}) \to 0$ as $n \to \infty$.

Likewise, it follows that $\lim_{n\to\infty} P(X_{[1]} = X_{[2]}) = \lim_{n\to\infty} P(X_{[1]} = X_{[3]}) = 0$. Hence $P(E_2(n)) \to 0$ as $n \to \infty$. Thus $P(E_3(n)) \to 1$ as $n \to \infty$. □

As a Corollary to Theorem 4.23, we have

Corollary 4.24 *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. Then $P(E_3(n)) \to 1$ as $n \to \infty$.*

Theorem 4.25 *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. For $r = 3/2$, the expected area of the the Γ_1-region, $\mathbf{E}\left[A\left(\Gamma_1^r(\mathcal{X}_n, M_C)\right)\right]$, converges to zero, at rate $O\left(n^{-2}\right)$.*

Proof By the geometry invariance property, we can assume that $T(\mathcal{Y}_3)$ is the standard equilateral triangle T_e. For $r = 3/2$ and $M = M_C$, and sufficiently large n, we have $\Gamma_1^r\left(X_{[i]}, M_C\right) \cap R_{M_C}(\mathsf{y}_i)$ is a triangle for $i = 1, 2, 3$ w.p. 1. See Fig. 8. With the realization of the edge extrema denoted as $x_{[i]} = (x_i, y_i)$ close enough to e_i, for $i = 1, 2, 3$, $\Gamma_1^{3/2}\left(x_{[1]}, M_C\right) \cap R_{M_C}(\mathsf{y}_1)$ is the triangle with vertices

$$\left(\frac{\sqrt{3}}{3}y_2 + x_2 - \frac{1}{2}, -\frac{\sqrt{3}}{18}\left(-9 + 2\sqrt{3}\,y_2 + 6x_2\right)\right),\ \left(\frac{1}{2}, \frac{\sqrt{3}}{9}\left(-\frac{9}{2} + 2\sqrt{3}\,y_2 + 6x_2\right)\right),\ \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right),$$

$\Gamma_1^{3/2}\left(x_{[2]}, M_C\right) \cap R_{M_C}(\mathsf{y}_2)$ is the triangle with vertices

$$\left(\frac{1}{2}, \frac{\sqrt{3}}{18}\left(3 + 4\sqrt{3}\,y_3 - 12x_3\right)\right),\ \left(\frac{1}{2} - \frac{\sqrt{3}}{3}y_3 + x_3, -\frac{\sqrt{3}}{18}\left(-3 + 2\sqrt{3}\,y_3 - 6x_3\right)\right),\ \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right),$$

and $\Gamma_1^{3/2}\left(x_{[3]}, M_C\right) \cap R_{M_C}(\mathsf{y}_3)$ is the triangle with vertices

$$\left(-\frac{\sqrt{3}}{6}\left(-\sqrt{3} + 4y_1\right), \frac{\sqrt{3}}{6} + \frac{2}{3}y_1\right),\ \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right),\ \left(\frac{\sqrt{3}}{6}\left(\sqrt{3} + 4y_1\right), \frac{\sqrt{3}}{6} + \frac{2}{3}y_1\right).$$

Then for sufficiently large n,

$$A\left(\Gamma_1^{3/2}(\mathcal{X}_n, M)\right)$$
$$= \frac{\sqrt{3}}{27}\left(3x_2 - 3 + \sqrt{3}\,y_2\right)^2 + \frac{\sqrt{3}}{27}\left(-3x_3 + \sqrt{3}\,y_3\right)^2 + \frac{4\sqrt{3}}{9}y_1^2$$
$$= \frac{\sqrt{3}}{9}\left(3x_2^2 - 6x_2 + 2\sqrt{3}\,y_2 x_2 - 2\sqrt{3}\,y_2 + y_2^2 + 3 + y_3^2 - 2\sqrt{3}\,y_3 x_3 + 3x_3^2 + 4y_1^2\right).$$

To find the expected area, we need the joint density of the $X_{[i]}$. By Theorem 4.23, the edge extrema are all distinct with probability 1 as $n \to \infty$. Let $T(\zeta)$ be the triangle formed by the lines at $x_{[i]}$ parallel to e_i for $i = 1, 2, 3$ where $\zeta = (x_1, y_1, x_2, y_2, x_3, y_3)$. See Fig. 8 (right).

The asymptotically accurate joint pdf of $X_{[i]}$ is

$$f_3(\zeta) = n(n-1)(n-2)\left[\frac{A(T(\zeta))}{A(T(\mathcal{Y}_3))}\right]^{n-3}\frac{1}{A(T(\mathcal{Y}_3))^3}$$
$$= n(n-1)(n-2)\left[\sqrt{3}/36\left(-2\sqrt{3}\,y_1 + \sqrt{3}\,y_3 - 3x_3 + \sqrt{3}\,y_2 + 3x_2\right)^2\right]^{n-3}\Big/\left(\sqrt{3}/4\right)^n.$$

with the support $D_S = \left\{\zeta \in \mathbb{R}^6 : (x_i, y_i)\text{'s are distinct}\right\}$.

For sufficiently large n,

$$\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M\right)\right)\right] \approx \int A\left(\Gamma_1^{3/2}(\mathcal{X}_n, M_C)\right) f_3(\zeta)\,d\zeta$$
$$= \int A\left(\Gamma_1^{3/2}(\mathcal{X}_n, M_C)\right) n(n-1)(n-2)\left(\frac{A(T(\zeta))}{A(T(\mathcal{Y}_3))}\right)^{n-3}$$
$$\times \frac{1}{A(T(\mathcal{Y}_3))^3}\,d\zeta.$$

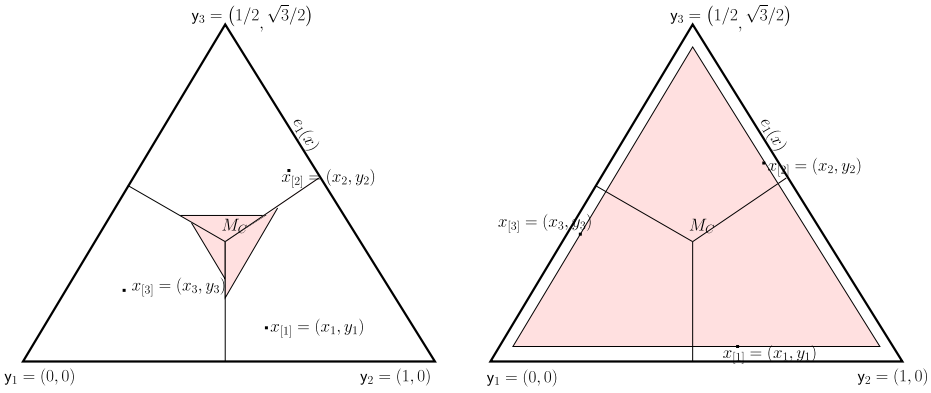

Fig. 8 The *shaded regions* are the triangular $\Gamma_1^{3/2}\left(X_{[i]}, M_C\right) \cap R_{M_C}(\mathsf{y}_i)$ regions for $i = 1, 2, 3$ (*left*). The figure for the joint pdf of $X_{[i]}$ where the shaded region is $T(\zeta)$ (*right*)

Let $G(\zeta) = A(T(\zeta))/A(T(\mathcal{Y}_3))$. Notice that the integrand is critical when $x_{[i]} \in e_i$ for $i = 1, 2, 3$, since $G(\zeta) = 1$ when $x_{[i]} \in e_i$ for each $i = 1, 2, 3$. So we make the change of variables $y_1 = z_1$, $y_2 = \sqrt{3}(1 - x_2) - z_2$, and $y_3 = \sqrt{3}x_3 - z_3$, then $G(\zeta)$ and $A(\Gamma_1(\mathcal{X}_n, N))$ become

$$G(\zeta) = G(z_1, z_2, z_3) = \left(2z_1 + z_3 - \sqrt{3} + z_2\right)^2 \Big/ 3 \text{ and}$$

$$A(\Gamma_1(\mathcal{X}_n, N)) = \sqrt{3}\left(z_2^2 + z_3^2 + 4z_1^2\right)/9,$$

respectively. Hence the integrand does not depend on x_1, x_2, x_3 and integrating with respect to x_1, x_2, x_3 yields a constant K. Now, the integrand is critical at $(z_1, z_2, z_3) = (0, 0, 0)$, since $G(0, 0, 0) = 1$. So let E_u^ε be the event that $0 \le z_i \le \varepsilon$ for $i = 1, 2, 3$ for $\varepsilon > 0$ small enough. Then making the change of variables $z_i = w_i/n$ for $i = 1, 2, 3$, we get $A\left(\Gamma_1^{3/2}(\mathcal{X}_n, M_C)\right) = O\left(n^{-2}\right)$ and $G(z_1, z_2, z_3)$ becomes $G(w_1, w_2, w_3) = 1 - \frac{1}{n}\left(2/\sqrt{3}(2w_1 + w_2 + w_3)\right) + O\left(n^{-2}\right)$, hence

$$\mathbf{E}\left[A\left(\Gamma_1^{3/2}(\mathcal{X}_n, M_C)\right)\right] \approx K \int_0^{n\varepsilon}\int_0^{n\varepsilon}\int_0^{n\varepsilon} A\left(\Gamma_1^{3/2}(\mathcal{X}_n, M_C)\right) n(n-1)(n-2) \times \frac{1}{n^3} G(w_1, w_2, w_3)^{n-3} dw_1 dw_2 dw_3,$$

letting $n \to \infty$

$$\approx K \int_0^{\infty}\int_0^{\infty}\int_0^{\infty} O\left(n^{-2}\right) \exp\left(-2/\sqrt{3}(2w_1 + w_2 + w_3)\right) dw_1 dw_2 dw_3 = O\left(n^{-2}\right),$$

since $\int_0^\infty \int_0^\infty \int_0^\infty \exp\left(-2/\sqrt{3}(2w_1 + w_2 + w_3)\right) dw_1 dw_2 dw_3 = 3\sqrt{3}/16$ which is a finite constant. Hence $\mathbf{E}\left[A\left(\Gamma_1^{3/2}(\mathcal{X}_n, M)\right)\right] \to 0$ as $n \to \infty$ at the rate $O\left(n^{-2}\right)$. □

In fact, in a similar fashion, one can demonstrate that the rate of convergence of the result in Theorem 4.20 is also $O(n^{-2})$ as $n \to \infty$.

In $\mathbb{R}^d$ with $d > 2$ and $\mathcal{Y}_{d+1} = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_{d+1}\}$, let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\mathfrak{S}(\mathcal{Y}_{d+1}))$. We can assume that $\mathfrak{S}(\mathcal{Y}_{d+1})$ to be the standard regular d-dimensional polytope with $d + 1$ vertices due to the geometry invariance. The region $\Gamma_1^r(\mathcal{X}_n, M)$ is determined by the (closest) face extrema $X_{[i]} \in \operatorname{argmin}_{X \in \mathcal{X}_n} d(X, \varphi_i)$, for $i \in \{1, 2, \ldots, (d+1)\}$. So, to find the expected volume of $\Gamma_1(\mathcal{X}_n, N, M)$, we need to find the expected locus of $X_{[i]}$; i.e., the expected distance of $X_{[i]}$ from φ_i. The results in Lemma 4.19 and Theorem 4.20 can be generalized with slight modifications as follows.

Lemma 4.26 *Let $D_i(n) := d\left(X_{[i]}, \varphi_i\right)$ for $i \in \{1, 2, \ldots, (d+1)\}$ and $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\mathfrak{S}(\mathcal{Y}_{d+1}))$. Then $\mathbf{E}[D_i(n)] \to 0$ (i.e., the expected locus of $X_{[i]}$ converges to face φ_i) for each $i \in \{1, 2, \ldots, (d+1)\}$, as $n \to \infty$.*

Theorem 4.27 *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\mathfrak{S}(\mathcal{Y}_{d+1}))$ and $M \in \mathfrak{S}(\mathcal{Y}_{d+1})^o$. Then $\mathbf{E}[\mu(\Gamma_1^r(\mathcal{X}_n, M))] \to \mu(\mathscr{R}_S(N_{PE}^r, M))$ as $n \to \infty$.*

We also conjecture that the rate of convergence in the above theorem is $O\left(n^{-d}\right)$.

5 The η-Values for PCDs

A problem of interest is finding, if possible, a subset of B, say G, such that $\Gamma_1(B, N) = \bigcap_{x\in G} \Gamma_1(x, N)$. This implies that only the points in G are *active* in determining $\Gamma_1(B, N)$.

Definition 5.1 An *active set* of points $S_A(B) \subseteq \Omega$ for determining $\Gamma_1(B, N)$ is defined to be a subset of B such that $\Gamma_1(B, N) = \bigcap_{x\in S_A(B)} \Gamma_1(x, N)$.

This definition allows B to be an active set, which always holds by Lemma 4.4. When B is a set of finitely many points, so is the associated active set. Among the active sets, we seek an active set of minimum cardinality.

Definition 5.2 Let B be a set of finitely many points. An active subset of $B \subset \Omega$ is called a *minimal active subset*, denoted $S_\mu(B)$, if there is no other active subset S_A of B such that $S_A(B) \subsetneq S_\mu(B)$. The minimum cardinality among the active subsets of B is called the *η-value* and denoted as $\eta(B, N)$. An active subset of cardinality $\eta(B, N)$ is called a *minimum active subset* denoted as $S_M(B)$; that is, $\eta(B, N) := |S_M(B)|$.

Then for any proximity map N and $\mathcal{X}_n$, it follows trivially that $\eta_n(N) \leq n$, since

$$\eta_n(N) := \min_{A\subseteq\mathcal{X}_n} \left\{ |A| : \Gamma_1(\mathcal{X}_n, N) = \bigcap_{Z\in A} \Gamma_1(Z, N) \right\}.$$

Note that the Definitions 5.1 and 5.2 can be extended for any subset $B \subseteq \Omega$. Moreover, a minimal active set of minimum cardinality is a minimum active set. We will suppress the dependence on B for $S_A(B)$, $S_\mu(B)$, and $S_M(B)$ if there is no ambiguity. In particular, if $B = \mathcal{X}_n$ is a set of Ω-valued random variables, then S_A and S_M are random sets and $\eta_n(N)$ is a random quantity.

For example, in $\mathbb{R}$ with $\mathcal{Y}_2 = \{0, 1\}$ and $\mathcal{X}_n$ a random sample of size $n > 1$ from F with support in $(0, 1)$, we have $\Gamma_1(\mathcal{X}_n, N_S) = (X_{n:n}/2, (1 + X_{1:n})/2)$. So the extrema (minimum and maximum) of the set $\mathcal{X}_n$ are sufficient to determine the Γ_1-region; i.e., $S_M = \{X_{1:n}, X_{n:n}\}$. Then $\eta_n(N_S) = 1 + \mathbf{I}(n > 1)$ a.s. for $\mathcal{X}_n$ being a random sample from a continuous distribution with support in $(0, 1)$.

In the multidimensional case there is no natural extension of ordering that yields natural extrema such as minimum or maximum. To get the minimum active sets associated with our proximity maps, we will resort to some other type of extrema, such as, the closest points to edges (i.e., edge extrema) or vertices in $T(\mathcal{Y}_3)$.

Proposition 5.3 *Let $\mathcal{X}_n$ be a random sample from a continuous distribution F on $T(\mathcal{Y}_3)$ and let $\eta_n(N^r_{PE})$ be the η-value associated with the PCD based on N^r_{PE} and $\mathcal{X}_n$. For proportional-edge proximity maps with M-vertex regions, we have $\eta_n(N^r_{PE}) \leq 3$ with equality holding with positive probability for $n \geq 3$.*

Proof From Theorem 4.16, it follows that $\eta_n(N^r_{PE}) \leq 3$. Furthermore, $X_{[i]}$ is unique for each edge e_i a.s. since F is continuous and there are three distinct edge extrema with positive probability. Hence $P(\eta_n(N^r_{PE}) = 3) > 0$ for $n \geq 3$. □

Recall that if $M \in (\mathscr{T}^r)^o$ with $1 \le r < 3/2$, then the region $\Gamma_1^r(\mathcal{X}_n, M)$ will be empty with probability 1 as $n \to \infty$. In such a case, there is no Γ_1-region to construct. But the definition of the η-value still works in the sense that $\Gamma_1^r(\mathcal{X}_n, M) = \emptyset = \bigcap_{x \in S_M(\mathcal{X}_n)} \Gamma_1^r(x, M)$ (see Definition 5.1 for S_M) and $\Gamma_1^r(x, M) \neq \emptyset$ for all $x \in \mathcal{X}_n$ since $x \in \Gamma_1^r(x, M)$.

Note that $P(\eta_n(N_{PE}^r) = 3) \to 1$ as $n \to \infty$ for $\mathcal{X}_n$ a random sample from F with positive density around the vertices $\mathcal{Y}_3$, since edge extrema are distinct with probability 1 as $n \to \infty$ as shown in Theorem 4.23.

Theorem 4.23 also implies that for $\mathcal{X}_n$ from an F with support as described in the theorem, the asymptotic distribution of $\eta_n(N_{PE}^r)$ is degenerate with $P(\eta_n(N_{PE}^r) = 3) \to 1$ as $n \to \infty$. Likewise, the same result holds for $\mathcal{X}_n$ from uniform data in $T(\mathcal{Y}_3)$. But for finite n, we have $\eta_n(N_{PE}^r)$ for $X_i \overset{iid}{\sim} F$ where F is as in Theorem 4.23, has the following non-degenerate distribution.

$$\eta_n(N_{PE}^r) = \begin{cases} 2 & \text{wp } \pi_2(n) \\ 3 & \text{wp } \pi_3(n) = 1 - \pi_2(n), \end{cases}$$

where $\pi_2(n) \in (0, 1)$ is the probability of edge extrema for any two distinct edges being concurrent.

Remark 5.4 If $\mathcal{X}_n$ is a random sample from F such that $\mathcal{S}(F) \cap \{x \in T(\mathcal{Y}_3) : d(x, e_i) \le \varepsilon_1\}$ has positive measure and $\mathcal{S}(F) \cap B(\mathsf{y}_i, \varepsilon_2) = \emptyset$ for some $\varepsilon_1, \varepsilon_2 > 0$ for each $i = 1, 2, 3$, then $P(E_3(n)) \to 1$ as $n \to \infty$ follows trivially.

Note that, for $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$, we have $P(\eta_n(N_{PE}^r) = 3) \to 1$ as $n \to \infty$, since the edge extrema are distinct with probability 1 as $n \to \infty$. Note also that $\eta_n(N_{PE}^{r_1}) \overset{d}{=} \eta_n(N_{PE}^{r_2})$ for all $(r_1, r_2) \in [1, \infty) \times [1, \infty)$, where $\overset{d}{=}$ stands for "equality in distribution".

We have shown that the upper bound for N_{PE}^r exists: $\eta_n(N_{PE}^r) \le 3$. However, for arc-slice PCDs, finding a $k < n$ such that $\eta_n(N_{AS}) \le k$ for all n is still an open problem. The same holds for $\eta_n(N_S)$.

Below we state a condition for $N(\cdot, M)$ defined with M-vertex regions to have $\eta_n(N) \le 3$ for $\mathcal{X}_n$ with support in $T(\mathcal{Y}_3)$.

Theorem 5.5 *Suppose $N(\cdot, M)$ is a proximity region defined with M-vertex regions and B is a set of n distinct points in $T(\mathcal{Y}_3)$. Then $\eta(B, N) \le 3$ if*

(i) *for each $\mathsf{y}_i \in \mathcal{Y}_3$ there exists a point $x(\mathsf{y}_i) \in B$ (i.e., related to y_i) such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \Gamma_1(x(\mathsf{y}_i), N) \cap R_M(\mathsf{y}_i)$,*
or

(ii) *there exist points $x(\mathsf{y}_j), x(\mathsf{y}_k) \in B$ such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \Gamma_1(x(\mathsf{y}_j), N) \cap \Gamma_1(x(\mathsf{y}_k), N) \cap R_M(\mathsf{y}_i)$ for $j, k \neq i$ with $i \in \{1, 2, 3\}$ and $(j, k) \in \{(1, 2), (1, 3), (2, 3)\}$.*

Proof Let $B = \{x_1, x_2, \ldots, x_n\} \subset T(\mathcal{Y}_3)$.

(i) Suppose there exists a point $x(\mathsf{y}_i) \in B$ such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \Gamma_1(x(\mathsf{y}_i), N) \cap R_M(\mathsf{y}_i)$ for each $i \in \{1, 2, 3\}$. Then

$$\begin{aligned}\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) &= \Gamma_1(x(\mathsf{y}_i), N) \cap R_M(\mathsf{y}_i) \\ &= \bigcap_{i=1}^{n} \left[\Gamma_1(x_i, N) \cap R_M(\mathsf{y}_i)\right] = \bigcap_{q=1}^{3} \left[\Gamma_1\left(x(\mathsf{y}_q), N\right) \cap R_M(\mathsf{y}_i)\right] \\ &= \left[\bigcap_{q=1}^{3} \Gamma_1\left(x(\mathsf{y}_q), N\right)\right] \cap R_M(\mathsf{y}_i)\end{aligned}$$

and

$$\Gamma_1(B, N) = \bigcup_{i=1}^{3} \left[\Gamma_1(B, N) \cap R_M(\mathsf{y}_i)\right] = \bigcup_{i=1}^{3} \left(\left[\bigcap_{q=1}^{3} \Gamma_1\left(x(\mathsf{y}_q), N\right)\right] \cap R_M(\mathsf{y}_i)\right).$$

Hence, we get

$$\Gamma_1(B, N) = \bigcap_{i=1}^{3} \Gamma_1(x(\mathsf{y}_i), N).$$

Thus, the minimum active set $S_M \subseteq \{x(\mathsf{y}_1), x(\mathsf{y}_2), x(\mathsf{y}_3)\}$, which implies $\eta(B, N) \leq 3$. The η-value $\eta(B, N) < 3$ will hold if $x(\mathsf{y}_i)$ are not all distinct.

(ii) Suppose there exist points $x(\mathsf{y}_j)$ and $x(\mathsf{y}_k)$ such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \left[\Gamma_1\left(x(\mathsf{y}_j), N\right) \cap \Gamma_1(x(\mathsf{y}_k), N)\right] \cap R_M(\mathsf{y}_i)$ for $j, k \neq i$. Then

$$\begin{aligned}\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) &= \left[\Gamma_1\left(x(\mathsf{y}_j), N\right) \cap \Gamma_1(x(\mathsf{y}_k), N)\right] \cap R_M(\mathsf{y}_i) \\ &= \bigcap_{i=1}^{n} \left[\Gamma_1(x_i, N) \cap R_M(\mathsf{y}_i)\right] = \bigcap_{q=1}^{3} \left[\Gamma_1\left(x(\mathsf{y}_q), N\right) \cap R_M(\mathsf{y}_i)\right] \\ &= \left[\bigcap_{q=1}^{3} \Gamma_1\left(x(\mathsf{y}_q), N\right)\right] \cap R_M(\mathsf{y}_i)\end{aligned}$$

and

$$\Gamma_1(B, N) = \bigcup_{i=1}^{3} \left[\Gamma_1(B, N) \cap R_M(\mathsf{y}_i)\right] = \bigcup_{i=1}^{3} \left(\left[\bigcap_{q=1}^{3} \Gamma_1\left(x(\mathsf{y}_q), N\right)\right] \cap R_M(\mathsf{y}_i)\right).$$

Hence, we get $\Gamma_1(B, N) = \bigcap_{i=1}^{3} \Gamma_1(x(\mathsf{y}_i), N)$. Therefore, the minimum active set $S_M \subseteq \{x(\mathsf{y}_1), x(\mathsf{y}_2), x(\mathsf{y}_3)\}$ which implies $\eta(B, N) \leq 3$. □

Notice that N_{PE}^r satisfies condition (i) of Theorem 5.5.

Below we state some conditions for $N(\cdot, M)$ defined with M-edge regions to have η-value less than equal to 3. See Ceyhan et al. (2007) for an example of proximity region which is constructed by using M-edge regions.

Theorem 5.6 *Suppose $N(\cdot, M)$ is a proximity region defined with M-edge regions and B is set of n distinct points in $T(\mathcal{Y}_3)$. Then $\eta(B, N) \leq 3$ if*

(i) *for each $e_i \in \{e_1, e_2, e_3\}$, there exists a point $x(e_i) \in B$ such that $\Gamma_1(B, N) \cap R_M(e_i) = \Gamma_1(x(e_i), N) \cap R_M(e_i)$,*
or

(ii) *there exist points $x(e_j), x(e_k) \in B$ such that $\Gamma_1(B, N) \cap R_M(e_i) = [\Gamma_1(x(e_j), N) \cap \Gamma_1(x(e_k), N)] \cap R_M(e_i)$ for $j, k \neq i$ with $i \in \{1, 2, 3\}$ and $(j, k) \in \{(1, 2), (1, 3), (2, 3)\}$.*

The proof of this theorem is same as the proof of Theorem 5.5 with vertices being replaced by edges.

In $\mathbb{R}^d$ with $d > 2$ the results in Proposition 5.3 and Theorem 4.23 can be extended as follows:

Proposition 5.7 *Let $\mathcal{X}_n$ be a random sample from a continuous distribution F on $\mathfrak{S}(\mathcal{Y}_{d+1})$ and let $\eta_n(N^r_{PE})$ be the η-value associated with the PCD based on N^r_{PE} and $\mathcal{X}_n$. For proportional-edge proximity maps with M-vertex regions, $\eta_n(N^r_{PE}) \leq d + 1$ with equality holding with positive probability for $n \geq d + 1$.*

Theorem 5.8 *Let $\mathcal{X}_n$ be a random sample from F such that $B(\mathsf{y}_i, \varepsilon) \subseteq \mathcal{S}(F)$ for some $\varepsilon > 0$ and for each $i = 1, 2, \ldots, (d + 1)$, and let $E_{d+1}(n)$ be the event that (closest) face extrema are distinct. Then $P(E_{d+1}(n)) \to 1$ as $n \to \infty$.*

The other results can be extended similarly.

6 The κ-Values for PCDs

Recall that the domination number, $\gamma_n(N)$ is the cardinality of a minimum dominating set of the PCD based on N. So by definition, $\gamma_n(N) \leq n$. We seek an a.s. least upper bound for $\gamma_n(N)$.

Definition 6.1 (κ-Value) Let $\mathcal{X}_n$ be a random sample from F on $T(\mathcal{Y}_3)$ and let $\gamma_n(N)$ be the domination number of the PCD based on a proximity map N and $\mathcal{X}_n$. The general a.s. least upper bound for $\gamma_n(N)$ that works for all $n \geq 1$ is called the *κ-value*; i.e., $\kappa_n(N) := \min\{k(n) : \gamma(\mathcal{X}_n, N) \leq k(n)$ a.s. for all $n \geq 1\}$.

It is more desirable to have a κ-value that is independent of n. Further, if $\kappa_n(N) = \kappa$ exists for the PCD based on the proximity map $N(\cdot)$ and is independent of n, then the domination number has the following discrete probability mass function: $P(\gamma(\mathcal{X}_n, N) = k) = p_k$ for $k = 1, 2, \ldots, \kappa$.

In $\mathbb{R}$ with $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\}$, for $\mathcal{X}_n$ a random sample from F with density f being positive around y_1 and y_2, we have $\gamma_n(N_S) \leq 2$ with equality holding with positive probability. Hence $\kappa_n(N_S) = 2$. The same holds for N_{AS} also. But in $\mathbb{R}^d$ with $d > 1$, finding $\kappa_n(N)$ for $N \in \{N_S, N_{AS}\}$ is an open problem. Next, we investigate the κ-values for the PCD based on N^r_{PE} in $\mathbb{R}^2$.

Theorem 6.2 *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$. Then for the PCD based on $N^r_{PE}(\cdot, M)$ and $\mathcal{X}_n$, we have $\kappa_n\left(N^r_{PE}\right) = 3$.*

Proof For $N^r_{PE}(\cdot, M)$, pick the point in vertex region $R_M(\mathsf{y}_i)$ closest to edge e_i; that is, pick $U_i \in \operatorname{argmin}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_i)} d(X, e_i) = \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_i)} d(\ell(\mathsf{y}, X), \mathsf{y}_i)$ in the vertex region for which $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \neq \emptyset$ for $i \in \{1, 2, 3\}$. Note that as $n \to \infty$, we have $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \neq \emptyset$ for all $i \in \{1, 2, 3\}$ a.s., and also U_i is unique a.s. for each i, since X is from $\mathcal{U}(T(\mathcal{Y}_3))$. Then $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \subset N^r_{PE}(U_i, M)$. Hence $\mathcal{X}_n \subset \bigcup_{i=1}^3 N^r_{PE}(U_i, M)$. So $\gamma_n\left(N^r_{PE}, M_C\right) \leq 3$ with equality holding with positive probability. Thus $\kappa_n\left(N^r_{PE}\right) = 3$. □

One property of proximity maps that makes $\kappa_n(N) < n$ is that the probability of having an $\mathcal{X}_n$ for which $N(X) \cap \mathcal{X}_n = \{X\}$ for all $X \in \mathcal{X}_n$ is zero.

Below we state a condition for $\kappa_n(N(\cdot, M)) \leq 3$ for $N(\cdot, M)$ defined with M-vertex regions.

Theorem 6.3 *Suppose $N(\cdot, M)$ is defined with M-vertex regions with $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$ and $N(x, M)$ gets larger as $d(\ell(\mathsf{y}, x), \mathsf{y})$ increases for $x \in R_M(\mathsf{y})$ in the sense that $N(x, M) \subseteq N(z, M)$ for all $x, z \in R_M(\mathsf{y})$ when $d(\ell(\mathsf{y}, x), \mathsf{y}) \leq d(\ell(\mathsf{y}, z), \mathsf{y})$. Furthermore, $N(X, M) \subsetneq N(Z, M)$ for all $X, Z \in R_M(\mathsf{y})$ occurs with positive probability when $d(\ell(\mathsf{y}, X), \mathsf{y}) < d(\ell(\mathsf{y}, Z), \mathsf{y})$ occurs with positive probability for X, Z from F. Then $\kappa_n(N) \leq 3$.*

Proof When $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \neq \emptyset$, pick one of the points $U_i(n) \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_i)} d(\ell(\mathsf{y}_i, X), \mathsf{y}_i)$, then $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \subset N(U_i(n))$ for each $i \in \{1, 2, 3\}$. So $\gamma(\mathcal{X}_n, N, M) \leq 3$, and hence $\kappa_n(N) \leq 3$. □

Notice that N^r_{PE} satisfies the conditions of Theorem 6.3.

Theorem 6.4 *Suppose $N(\cdot, M)$ is defined with M-edge regions and $N(x, M)$ gets larger as $d(x, e)$ increases for $x \in R_M(e)$ in the sense that $N(x, M) \subseteq N(y, M)$ for all $x, y \in R_M(e)$ when $d(x, e) \leq d(y, e)$. Suppose also that for X, Y from F, $N(X, M) \subsetneq N(Y, M)$ for all $X, Y \in R_M(e)$ occurs with positive probability when $d(X, e) < d(Y, e)$ occurs with positive probability. Then $\kappa_n(N) \leq 3$.*

Proof When $\mathcal{X}_n \cap R_M(e_i) \neq \emptyset$, pick one of the points $U_{[i]} \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(e_i)} d(X, e_i)$. Then $\mathcal{X}_n \cap R_M(e_i) \subset N(U_{[i]})$ for each $i \in \{1, 2, 3\}$. So $\gamma(\mathcal{X}_n, N, M) \leq 3$, and hence $\kappa_n(N) \leq 3$. □

In Theorems 6.3 and 6.4, we have an upper bound for $\kappa_n(N)$. To determine an exact value for $\kappa_n(N)$ we need further restrictions. Let $A_1 := \{x \in T(\mathcal{Y}_3) : \mu(N(x, M)) = \mu(T(\mathcal{Y}_3))\}$ and $A_2 := \{(x, y) \in T(\mathcal{Y}_3) \times T(\mathcal{Y}_3) : \mu(N(x, M) \cup N(y, M)) = \mu(T(\mathcal{Y}_3))\}$. If in addition to the hypothesis of Theorem 6.3 (and 6.4) we have A_1 and A_2 that have zero measure (e.g., zero area for continuous distributions with support in $T(\mathcal{Y}_3)$), then $\kappa_n(N) = 3$ would hold.

In $\mathbb{R}^d$ with $d > 2$, the result in Theorem 6.2 can be extended as follows:

Theorem 6.5 *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\mathfrak{S}(\mathcal{Y}_{d+1}))$ and $M \in \mathbb{R}^d \setminus \mathcal{Y}_{d+1}$. Then for the PCD based on $N^r_{PE}(\cdot, M)$ and $\mathcal{X}_n$, we have $\kappa_n\left(N^r_{PE}\right) = d + 1$.*

The other results can be extended similarly.

7 The Γ_k-Regions for Proximity Maps

We can also define the regions associated with $\gamma_n(N) = k$ for $k \leq n$. In $\mathbb{R}$ with $\mathcal{Y}_2 = \{0, 1\}$, it follows that $\gamma_n(N) \leq 2$, hence we can only define Γ_2-regions. Recall that $\gamma_n(N) = 2$ iff $\mathcal{X}_n \cap \left[\frac{x_{n:n}}{2}, \frac{1+x_{1:n}}{2}\right] = \emptyset$ iff $\mathcal{X}_n \subset [0, 1] \setminus \Gamma_1(\mathcal{X}_n, N)$. So

$$\begin{aligned}\Gamma_2(\mathcal{X}_n, N) &= \left\{(x, y) \in [0, 1]^2 :\ \mathcal{X}_n \subset N(x) \cup N(y);\ x, y \notin \Gamma_1(\mathcal{X}_n, N)\right\} \\ &= \left\{(x, y) \in ([0, 1] \setminus \Gamma_1(\mathcal{X}_n, N))^2 :\ \mathcal{X}_n \subset N(x) \cup N(y)\right\}\end{aligned}$$

where $A^2 = A \times A$. Notice that $\Gamma_2(\mathcal{X}_n, N) \subseteq [0, 1]^2$. Let $x_{[1]} := \operatorname{argmin}_{x \in \mathcal{X}_n \cap (0,1/2)}(1/2 - x)$ and $x_{[2]} := \operatorname{argmin}_{x \in \mathcal{X}_n \cap (1/2,1)}(x - 1/2)$, then $\gamma_n(N) = 2$ iff $x_{[1]}, x_{[2]} \notin \Gamma_1(\mathcal{X}_n, N)$. In such a case $\mathcal{X}_n \subset N\left(x_{[1]}\right) \cup N\left(x_{[2]}\right)$ by construction.

In general, Γ_k-regions can be defined as follows for $k > 1$.

Definition 7.1 (Γ_k-Region) The Γ_2-*region* for proximity map $N(\cdot)$ and set $B \subset \Omega$ is $\Gamma_2(B, N) = \{(x, y) \in [\Omega \setminus \Gamma_1(B)]^2 :\ B \subseteq N(x) \cup N(y)\}$. In general, Γ_k-*region* for proximity map $N(\cdot)$ and set $B \subset \Omega$ for $k = 1, 2, \ldots, n$ is

$$\begin{aligned}\Gamma_k(B, N) = \Bigg\{&(x_1, x_2, \ldots, x_k) \in \Omega^k : B \subseteq \bigcup_{i=1}^{k} N(x_i) \text{ and all possible } m\text{-permutations} \\ &(u_1, u_2, \ldots, u_m) \text{ of } (x_1, x_2, \ldots, x_k) \text{ satisfy } (u_1, u_2, \ldots, u_m) \notin \Gamma_m(B, N) \\ &\text{for each } m = 1, 2, \ldots, k-1\Bigg\}.\end{aligned}$$

Note that Γ_k-regions are defined for $k \leq n$ and they might be empty. Moreover, Γ_k-regions are in Ω^k. Let $\left\{{n \atop m}\right\}$ denote the Stirling partition number for a set of size n into m blocks and let $\left\{{A \atop m}\right\}$ denote all Stirling partitions of a set A into m blocks; that is,

$$\begin{aligned}\left\{{A \atop m}\right\} := \Bigg\{&\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_m\} :\ \mathcal{B}_i \neq \emptyset, \mathcal{B}_i \subset A, \mathcal{B}_i \cap \mathcal{B}_j = \emptyset \text{ for } i \neq j,\ i, j = 1, 2, \ldots, m; \\ &A = \bigcup_i \mathcal{B}_i\Bigg\}.\end{aligned}$$

In particular, for a set B of size n, the Stirling partition $\left\{{B \atop 2}\right\}$ is the unordered pair of blocks $\mathcal{B}_1$ and $\mathcal{B}_2$ such that $\mathcal{B}_i \neq \emptyset$ and $\mathcal{B}_i \subsetneq B$ for $i = 1, 2$, and $\mathcal{B}_1 \cup \mathcal{B}_2 = B$. Note that $\mathcal{B}_2 = B \setminus \mathcal{B}_1$. Then we have the following result.

Proposition 7.2 *Let B be a set of finitely many points in Ω and $\Gamma_2(B, N)$ be the corresponding Γ_2-region associated with the proximity map $N(\cdot)$. Then $\Gamma_2(B, N) = \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {B \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N)] \setminus \left[\Gamma_1(B, N)^2\right]$ for any Stirling blocks $\mathcal{B}_1$ and $\mathcal{B}_2$ in $\left\{ {B \atop 2} \right\}$.*

Proof Given B with $|B| = n$, suppose $(u, v) \in \Gamma_2(B, N)$, then $B \subset N(u) \cup N(v)$ and $u, v \notin \Gamma_1(B, N)$. Let $\mathcal{B}_1 = B \cap N(u)$ and $\mathcal{B}_2 = [B \cap N(v)] \setminus N(u)$. Then $\mathcal{B}_1$ and $\mathcal{B}_2$ are two Stirling blocks in $\left\{ {B \atop 2} \right\}$. Hence $\mathcal{B}_1 \subset N(u)$ implies that $u \in \Gamma_1(\mathcal{B}_1, N) \setminus \Gamma_1(B, N)$ and $\mathcal{B}_2 \subset N(v)$ implies that $v \in \Gamma_1(\mathcal{B}_2, N) \setminus \Gamma_1(B, N)$, hence $\Gamma_2(B, N) \subseteq \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {B \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N)] \setminus \left[\Gamma_1(B, N)^2\right]$. The other direction is trivial, hence the desired result follows. □

In $\mathbb{R}$ with $N = N_S$, we can exploit the natural ordering available.

Proposition 7.3 *In $\mathbb{R}$ with $\mathcal{Y}_2 = \{0, 1\}$, we have $\Gamma_2(\mathcal{X}_n, N_S) = \bigcup_{k=1}^{n-1} \left(\frac{x_{k:n}}{2}, \frac{x_{n:n}}{2}\right) \times \left(\frac{1+x_{1:n}}{2}, \frac{1+x_{(k+1):n}}{2}\right)$.*

Proof From Proposition 7.2, we have $\Gamma_2(\mathcal{X}_n, N_S) = \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {\mathcal{X}_n \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S)] \setminus (\Gamma_1(\mathcal{X}_n, N_S)^2)$. Let $\{\mathcal{B}_1, \mathcal{B}_2\}$ be a Stirling partition in $\left\{ {\mathcal{X}_n \atop 2} \right\}$. First observe that $\Pi_1(k) = \{x_{1:n}, x_{2:n}, \ldots, x_{k:n}\}$ and $\Pi_2(k) = \{x_{(k+1):n}, x_{(k+2):n}, \ldots, x_{n:n}\}$ forms a Stirling partition of $\mathcal{X}_n$, and

$$\Gamma_1(\Pi_1(k), N_S) = \left(\frac{x_{k:n}}{2}, \frac{1 + x_{1:n}}{2}\right), \ \Gamma_1(\Pi_2(k), N_S) = \left(\frac{x_{n:n}}{2}, \frac{1 + x_{(k+1):n}}{2}\right),$$

and

$$\Gamma_1(\mathcal{X}_n, N_S) = \left(\frac{x_{n:n}}{2}, \frac{1 + x_{1:n}}{2}\right).$$

Then we have

$$\Gamma_1(\Pi_1(k), N_S) \setminus \Gamma_1(\mathcal{X}_n, N_S) = \left(\frac{x_{k:n}}{2}, \frac{x_{n:n}}{2}\right)$$

and

$$\Gamma_1(\Pi_2(k), N_S) \setminus \Gamma_1(\mathcal{X}_n, N_S) = \left(\frac{1 + x_{1:n}}{2}, \frac{1 + x_{(k+1):n}}{2}\right).$$

Now if $x_{1:n}$ and $x_{n:n}$ are in one of the Stirling partitions, say $\mathcal{B}_1$, then $[\Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S)] \setminus (\Gamma_1(\mathcal{X}_n, N_S)^2)$ will be empty, since $\Gamma_1(\mathcal{B}_1, N_S) \setminus (\Gamma_1(\mathcal{X}_n, N_S))$ is

empty. Furthermore, if $\max(\mathcal{B}_1) < \min(\mathcal{B}_2)$, then $\mathcal{B}_1$ and $\mathcal{B}_2$ do not constitute a Stirling partition. So we only consider Stirling partitions of the form

$$\Lambda_k = \left\{ \{\mathcal{B}_1, \mathcal{B}_2\} \in \left\{ \begin{matrix} \mathcal{X}_n \\ 2 \end{matrix} \right\} : \; x_{1:n} \in \mathcal{B}_1, \; x_{n:n} \in \mathcal{B}_2, \; \max(\mathcal{B}_1) = x_{k:n}, \right.$$
$$\left. \min(\mathcal{B}_2) = x_{l:n}, \; l < k \right\}.$$

Then for $\{\mathcal{B}_1, \mathcal{B}_2\} \in \Lambda_k$, we have $\Gamma_1(\mathcal{B}_1, N_S) = \left(\frac{x_{k:n}}{2}, \frac{1+x_{1:n}}{2}\right)$ and $\Gamma_1(\mathcal{B}_2, N_S) = \left(\frac{x_{n:n}}{2}, \frac{1+x_{l:n}}{2}\right)$. Hence, $\Gamma_1(\mathcal{B}_1, N_S) = \Gamma_1(\Pi_1(k), N_S)$ and $\Gamma_1(\mathcal{B}_2, N_S) \subsetneq \Gamma_1(\Pi_2(k), N_S)$, since $l < k$. Therefore

$$\bigcup_{\{\mathcal{B}_1, \mathcal{B}_2\} \in \Lambda_k} \Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S) = \Gamma_1(\Pi_1(k), N_S) \times \Gamma_1(\Pi_2(k), N_S). \tag{4}$$

Moreover, we have

$$\bigcup_{\{\mathcal{B}_1, \mathcal{B}_2\} \in \left\{ \begin{matrix} \mathcal{X}_n \\ 2 \end{matrix} \right\}} [\Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S)] = \bigcup_{k=1}^{n-1} \left[\bigcup_{\{\mathcal{B}_1, \mathcal{B}_2\} \in \Lambda_k} \Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S) \right]$$
$$= \bigcup_{k=1}^{n-1} \left[\Gamma_1(\Pi_1(k), N_S) \times \Gamma_1(\Pi_2(k), N_S) \right].$$

Since $\bigcup_{k=1}^{n-1} \left[\Gamma_1(\Pi_1(k), N_S) \times \Gamma_1(\Pi_2(k), N_S) \right] = \left(\frac{x_{k:n}}{2}, \frac{1+x_{1:n}}{2}\right) \times \left(\frac{x_{n:n}}{2}, \frac{1+x_{(k+1):n}}{2}\right)$ we have

$$\Gamma_2(\mathcal{X}_n, N_S) = \bigcup_{\{\mathcal{B}_1, \mathcal{B}_2\} \in \left\{ \begin{matrix} \mathcal{X}_n \\ 2 \end{matrix} \right\}} [\Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S)] \setminus (\Gamma_1(\mathcal{X}_n, N_S)^2)$$
$$= \bigcup_{k=1}^{n-1} \left[\Gamma_1(\Pi_1(k), N_S) \times \Gamma_1(\Pi_2(k), N_S) \right] \setminus (\Gamma_1(\mathcal{X}_n, N_S)^2$$
$$= \bigcup_{k=1}^{n-1} \left(\frac{x_{k:n}}{2}, \frac{x_{n:n}}{2} \right) \times \left(\frac{1 + x_{1:n}}{2}, \frac{1 + x_{(k+1):n}}{2} \right).$$

So we have the desired result. □

For N_{PE}^r, the Γ_1-region for blocks in Proposition 7.2, the region $\Gamma_1^r(\mathcal{B}_i, M)$ is determined by the edge extrema in $\mathcal{B}_i$ for $i = 1, 2$. But for N_{PE}^r, if $(u, v) \in \Gamma_2\left(\mathcal{X}_n, N_{PE}^r, M\right)$, then $(u, v) \notin R_M(\mathsf{y})^2$, since either $N_{PE}^r(u) \subseteq N_{PE}^r(v)$ or $N_{PE}^r(v) \subseteq N_{PE}^r(u)$ should hold if $(u, v) \in R_M(\mathsf{y})^2$.

For any proximity map N,

$$
\begin{aligned}
&P(\gamma_n(N) = 2) \\
&= P(\mathcal{X}_n^2 \cap \Gamma_2(\mathcal{X}_n, N) \neq \emptyset,\ \gamma_n(N) \neq 1) \\
&= P\left(\mathcal{X}_n^2 \bigcap \left[\bigcup_{\{\mathcal{B}_1, \mathcal{B}_2\} \in \binom{\mathcal{X}_n}{2}} \left[\Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N) \right] \setminus (\Gamma_1(\mathcal{X}_n, N))^2 \right] \neq \emptyset, \right. \\
&\qquad \left. \mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N) = \emptyset \right).
\end{aligned}
$$

A more compact way to write this is as $P(\gamma_n(N) > 2) = P\left(\mathcal{X}_n^2 \cap \Gamma_{\leq 2}(\mathcal{X}_n, N) = \emptyset\right)$ where $\Gamma_{\leq 2}(\mathcal{X}_n, N) = \bigcup_{\{\mathcal{B}_1, \mathcal{B}_2\} \in \binom{\mathcal{X}_n}{2}} \Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N)$.

Furthermore, for $k \geq 3$, the Γ_k-regions are defined similarly as

$$
\Gamma_k(\mathcal{X}_n, N) = \bigcup_{\{\mathcal{B}_1, \ldots, \mathcal{B}_k\} \in \binom{\mathcal{X}_n}{k}} \left[\Gamma_1(\mathcal{B}_1, N) \times \ldots \times \Gamma_1(\mathcal{B}_k, N) \right] \setminus \Gamma_1(\mathcal{X}_n, N)^k .
$$

Hence,

$$
\begin{aligned}
&P(\gamma_n(N) = k) \\
&= P\left(\mathcal{X}_n^k \cap \Gamma_k(\mathcal{X}_n, N) \neq \emptyset, \gamma_n(N) > k - 1\right) \\
&= P\left(\mathcal{X}_n^k \cap \left[\bigcup_{\{\mathcal{B}_1, \ldots, \mathcal{B}_k\} \in \binom{\mathcal{X}_n}{k}} \Gamma_1(\mathcal{B}_1, N) \ldots \times \Gamma_1(\mathcal{B}_k, N) \setminus \Gamma_1(\mathcal{X}_n, N)^k \right] \neq \emptyset \right).
\end{aligned}
$$

A more compact way to write this is as $P(\gamma_n(N) > k) = P\left(\mathcal{X}_n^k \cap \Gamma_{\leq k}(\mathcal{X}_n, N) \neq \emptyset\right)$ where $\Gamma_{\leq k}(\mathcal{X}_n, N) = \bigcup_{\{\mathcal{B}_1, \ldots, \mathcal{B}_k\} \in \binom{\mathcal{X}_n}{k}} \Gamma_1(\mathcal{B}_1, N) \times \ldots \times \Gamma_1(\mathcal{B}_k, N)$.

8 Discussion and Conclusions

In this article, we introduce new graph invariants associated with the domination number of proximity catch digraphs (PCDs) based on proximity maps and investigate their probabilistic behavior. Although Γ_1-regions and superset regions were introduced before (see Ceyhan and Priebe 2005; Ceyhan et al. 2006; Ceyhan and Priebe 2007), a thorough investigation is only performed in this article. A Γ_1-region is sort of a "dual" of the underlying proximity region and is closely associated with domination number being equal to one. We investigate the probabilistic behavior of Γ_1-regions for general proximity maps N and data points from a fairly general distribution F. We also extend this concept by introducing Γ_k-regions, which are associated with domination number being equal to k.

We introduce the quantities related to domination number, namely, η- and κ-values. The η-value is the minimum number of points in a set required to determine

the Γ_1-region for that set. We determine some general conditions that make $\eta_n(N) \leq 3$ for data in the triangle $T(\mathcal{Y}_3)$. The κ-value is the a.s. least upper bound for the domination number of the PCDs. We also determine some general conditions that make $\kappa_n(N) \leq 3$ for data in the triangle $T(\mathcal{Y}_3)$. We also extend the results to higher dimensions.

We provide a few PCD families, namely spherical, arc-slice (Ceyhan and Priebe 2003), and proportional-edge PCDs (Ceyhan et al. 2006), as illustrative examples. We discuss the construction of proximity regions and Γ_1-regions for these PCDs. Furthermore, we calculate the limit of the expected measure of Γ_1-regions for the spherical PCDs in $\mathbb{R}$ and proportional-edge PCDs for $\mathbb{R}^2$. Determining Γ_1-regions, $\eta-$ and κ-values for spherical and arc-slice PCDs contain many open problems and are subjects of ongoing research.

With the tools presented in this article, given a PCD, we can determine how it behaves in terms of domination number and the related invariants. For example, we can determine a.s. least upper bounds for the domination number of the new PCD.

Acknowledgements I would like to thank the anonymous referees, whose constructive comments and suggestions greatly improved the presentation and flow of this article. This research was supported by the research agency TUBITAK via the Kariyer Project # 107T647.

References

Billingsley P (1995) Probability and measure. Wiley-Interscience, New York

Boland PJ, Singh H, Cukic B (2004) The stochastic precedence ordering with applications in sampling and testing. J Appl Probab 41(1):73–82

Ceyhan E (2005) An investigation of proximity catch digraphs in Delaunay tessellations, also available as technical monograph titled "Proximity catch digraphs: auxiliary tools, properties, and applications" by VDM Verlag. ISBN: 978-3-639-19063-2. PhD thesis, The Johns Hopkins University, Baltimore, MD, 21218

Ceyhan E (2008) The distribution of the domination number of class cover catch digraphs for non-uniform one-dimensional data. Discrete Math 308:5376–5393

Ceyhan E (2009a) Extension of one-dimensional proximity regions to higher dimensions. arXiv:0902.1306v1 [math.MG]. Technical report # KU-EC-09-2, Koç University, Istanbul, Turkey

Ceyhan E (2009b) A probabilistic characterization of random proximity catch digraphs and the associated tools. arXiv:0903.5005v1 [math.PR]. Technical report # KU-EC-09-3, Koç University, Istanbul, Turkey

Ceyhan E (2010) Spatial clustering tests based on domination number of a new random digraph family. Commun Stat, Theory Methods. doi:10.1080/03610921003597211

Ceyhan E, Priebe C (2003) Central similarity proximity maps in Delaunay tessellations. In: Proceedings of the joint statistical meeting, statistical computing section, American Statistical Association

Ceyhan E, Priebe CE (2005) The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association. Stat Probab Lett 73:37–50

Ceyhan E, Priebe CE (2007) On the distribution of the domination number of a new family of parametrized random digraphs. Mod Assist Stat Appl 1(4):231–255

Ceyhan E, Priebe CE, Wierman JC (2006) Relative density of the random r-factor proximity catch digraphs for testing spatial patterns of segregation and association. Comput Stat Data Anal 50(8):1925–1964

Ceyhan E, Priebe CE, Marchette DJ (2007) A new family of random graphs for testing spatial segregation. Can J Stat 35(1):27–50

DeVinney J (2003) The class cover problem and its applications in pattern recognition. PhD thesis, The Johns Hopkins University, Baltimore, MD, 21218

DeVinney J, Priebe CE (2006) A new family of proximity graphs: class cover catch digraphs. Discrete Appl Math 154(14):1975–1982
DeVinney J, Wierman JC (2003) A SLLN for a one-dimensional class cover problem. Stat Probab Lett 59(4):425–435
DeVinney J, Priebe CE, Marchette DJ, Socolinsky D (2002) Random walks and catch digraphs in classification. In: Proceedings of the 34th symposium on the interface: computing science and statistics, vol 34. http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf
Devroye L, Gyorfi L, Lugosi G (1996) A probabilistic theory of pattern recognition. Springer, New York
Garfinkel RS, Nemhauser GL (1972) Integer programming. Wiley, New York
Janson S, Łuczak T, Ruciński A (2000) Random graphs. In: Wiley-Interscience series in discrete mathematics and optimization. Wiley, New York
Jaromczyk JW, Toussaint GT (1992) Relative neighborhood graphs and their relatives. Proc IEEE 80:1502–1517
Karr AF (1992) Probability. Springer, New York
Kimberling C (2008) Encyclopedia of triangle centers. http://faculty.evansville.edu/ck6/encyclopedia/ETC.html
Lee C (1998) Domination in digraphs. J Korean Math Soc 4:843–853
Marchette DJ, Priebe CE (2003) Characterizing the scale dimension of a high dimensional classification problem. Pattern Recogn 36(1):45–60
Okabe A, Boots B, Sugihara K, Chiu SN (2000) Spatial tessellations: concepts and applications of Voronoi diagrams. Wiley, New York
Paterson MS, Yao FF (1992) On nearest neighbor graphs. In: Proceedings of 19th int. coll. automata, languages and programming. Springer LNCS, vol 623, pp 416–426
Priebe CE, DeVinney JG, Marchette DJ (2001) On the distribution of the domination number of random class cover catch digraphs. Stat Probab Lett 55:239–246
Priebe CE, Marchette DJ, DeVinney J, Socolinsky D (2003a) Classification using class cover catch digraphs. J Classif 20(1):3–23
Priebe CE, Solka JL, Marchette DJ, Clark BT (2003b) Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays. Comput Stat Data Anal Visual 43-4:621–632
Sen M, Das S, Roy A, West D (1989) Interval digraphs: an analogue of interval graphs. J Graph Theory 13:189–202
Toussaint GT (1980) The relative neighborhood graph of a finite planar set. Pattern Recogn 12(4):261–268
Tuza Z (1994) Inequalities for minimal covering sets in sets in set systems of given rank. Discrete Appl Math 51:187–195
Wierman JC, Xiang P (2008) A general SLLN for the one-dimensional class cover problem. Stat Probab Lett 78(9):1110–1118
Xiang P, Wierman JC (2009) A CLT for a one-dimensional class cover problem. Stat Probab Lett 79(2):223–233