Statistical Methodology 33 (2016) 31–54

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Edge density of new graph types based on a random digraph family

Elvan Ceyhan

Department of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 31 March 2015
Received in revised form
3 April 2016
Accepted 14 July 2016
Available online 27 July 2016

MSC:
60D05
60F05
62E20
05C80
05C20
60C05

Keywords:
Arc density
Asymptotic normality
Central limit theorem
Delaunay tessellation and triangulation
Proximity catch digraph
Reflexivity
Underlying graph
U-statistic

ABSTRACT

We consider two types of graphs based on a family of proximity catch digraphs (PCDs) and study their edge density. In particular, the PCDs we use are a parameterized digraph family called proportional-edge (PE) PCDs and the two associated graph types are the "underlying graphs" and the newly introduced "reflexivity graphs" based on the PE-PCDs. These graphs are extensions of random geometric graphs where distance is replaced with a dissimilarity measure and the threshold is not fixed but depends on the location of the points. PCDs and the associated graphs are constructed based on data points from two classes, say $\mathcal{X}$ and $\mathcal{Y}$, where one class (say class $\mathcal{X}$) forms the vertices of the PCD and the Delaunay tessellation of the other class (i.e., class $\mathcal{Y}$) yields the (Delaunay) cells which serve as the support of class $\mathcal{X}$ points. We demonstrate that edge density of these graphs is a U-statistic, hence obtain the asymptotic normality of it for data from any distribution that satisfies mild regulatory conditions. The rate of convergence to asymptotic normality is sharper for the edge density of the reflexivity and underlying graphs compared to the arc density of the PE-PCDs. For uniform data in Euclidean plane where Delaunay cells are triangles, we demonstrate that the distribution of the edge density is geometry invariant (i.e., independent of the shape of the triangular support). We compute the explicit forms of the asymptotic normal distribution for uniform data in one Delaunay triangle in the Euclidean plane utilizing this geometry invariance property. We also provide various versions of edge density in the multiple triangle case. The approach presented here can also be extended for application to data in higher dimensions.

E-mail address: elceyhan@ku.edu.tr.

http://dx.doi.org/10.1016/j.stamet.2016.07.003
1572-3127/

1. Introduction

Proximity catch digraphs (PCDs) are a recently introduced digraph family and have applications in pattern classification and spatial data analysis. PCDs are random digraphs (i.e., directed graphs) in which each vertex corresponds to a data point, and arcs (i.e., directed edges) are defined in terms of some bivariate relation on the data. One type of PCD is the class cover catch digraph (CCCD) introduced by Priebe et al. [23] who gave the exact and the asymptotic distribution of its domination number for uniform data on bounded intervals in $\mathbb{R}$. Priebe et al. [25] and DeVinney and Priebe [13] applied the concept in higher dimensions and demonstrated relatively good performance of it in classification. Their methods involve *data reduction* (i.e., *condensing*) by using approximate minimum dominating sets as *prototype sets*, since finding the exact minimum dominating set is an NP-hard problem in general – e.g., for CCCDs in multiple dimensions – (see DeVinney and Priebe [13]). Our PCDs are constructed in a two-class setting with points from the class of interest (i.e., target class) that constitutes the vertices of the digraph. Let $\mathcal{X}_n$ and $\mathcal{Y}_m$ be two data sets of size n and m from classes $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let class $\mathcal{X}$ be the target class. Then the vertices of the PCD are $\mathcal{X}_n$ and there is an arc from $x_1 \in \mathcal{X}_n$ to $x_2 \in \mathcal{X}_n$, based on a binary relation which measures the relative proximity of x_1 and x_2 with respect to the $\mathcal{Y}$ points. This relative proximity is determined based on the Delaunay tessellation of $\mathcal{Y}$ points. The PCDs are also closely related to the class cover problem of Cannon and Cowen [5] where the goal is finding a cover for the target class (i.e., finding a set of regions that contain all the points from the target class). Ceyhan and Priebe [8] introduced a digraph family called *proportional-edge PCD* (PE-PCD) and calculated the asymptotic distribution of its arc density and used it in spatial pattern testing [10]. PE-PCDs are parameterized digraphs with an expansion and a centrality parameter.

We consider two graph types defined based on the digraphs (in particular on PE-PCDs). The *underlying graph* of a digraph is obtained when any arc (between two vertices) is replaced by an edge disallowing multi-edges [11]. We introduce another graph type by replacing each symmetric or reflexive arc (between two vertices) with an edge and removing the (nonsymmetric) single arcs, and call it *reflexivity graph*. In a digraph $D = (V, A)$ with vertex set V and arc set A, an arc (a, b) is *symmetric* iff $\{(a, b), (b, a)\} \subset A$, i.e., the points a and b satisfy *reflexivity* with respect to the binary relation defining the arcs. The reflexivity and underlying graphs based on the PE-PCDs are also generalized versions of random geometric graphs (RGGs) (see, e.g., Penrose [22] for an extensive treatment of RGGs). Applications of RGGs include modeling/understanding disease spread among trees scattered in a forest, communication between a set of nests of animals or birds in a region of interest or between stations in a country or nerve cells in a living organism. RGGs are based on vertices independently and identically (iid) generated in $\mathbb{R}^k$ and an edge is inserted between two vertices if the distance between them does not exceed a certain threshold value. On the other hand, in our graphs, the regions that determine the edges would depend on the vertices, hence the threshold is adjusted based on the location of the vertices; and instead of a distance, a dissimilarity measure is employed. PCDs might also be applied in similar settings as those of RGGs, e.g., in testing spatial interaction (as in [10,9]); and in similar settings as those of the CCCDs in pattern classification (as in [24]).

We investigate the properties of the graph invariant, called *edge density*, of the reflexivity and underlying graphs of PCDs. Edge density of a graph is the ratio of number of edges to the total number of edges possible with the same set of vertices. Hence for a graph $G_n = (V, E)$ with vertex set of size $|V| = n$ and edge set E, edge density is $2|E|/(n(n-1))$. The maximal density is 1, which is attained for complete graphs, while the minimal density is 0, which is attained when $E = \emptyset$. The average degree of the graph G_n is defined as $2|E|/n$, is closely related to edge density; it is simply a scaled version of edge density. Edge density is also defined as $|E|/n$ by Grünbaum [14] who studies it for 4-critical planar graphs. In this article, we only use the quantity $2|E|/(n(n-1))$ as the edge density for the graph $G_n = (V, E)$. Arc density of a digraph $\mathbf{D}_n = (V, A)$ with $|V| = n$ is defined similarly as $|A|/(n(n-1))$. Edge density is also instrumental in determining the density of higher order structures in graphs, e.g., minimal density of triangles in graphs is provided in terms of edge density by Razborov [26]. Michael [21] studies the edge density for another special type of geometric graph called sphere of influence graph. A local version of edge density is also defined and investigated for subgraphs [20]. Furthermore, Darst et al. [12] define another local version called internal edge

density to determine communities (i.e., clusters) in networks, and show that their method is capable of determining overlapping communities in both graphs and multi-graphs. Arc density also has uses in network data, e.g., in large social networks. For example, Boullé [3] uses a Bayesian approach to summarize the structure of a large directed multi-graph, based on non-parametric estimation of arc density using a piecewise constant estimation of the density of the arcs across the clusters, thereby automatically and reliably inferring the number of clusters.

Properly scaled, we demonstrate that the edge density of the reflexivity and underlying graphs of PE-PCDs is a U-statistic, which has asymptotic normality by the general central limit theory of U-statistics [15]. Furthermore, we derive the explicit form of the asymptotic normal distribution of the edge density of these graphs based on uniform data in a bounded region in $\mathbb{R}^2$. In particular, this bounded region is taken to be the convex hull of class $\mathcal{Y}$ points, and the arcs between class $\mathcal{X}$ points are constructed using the Delaunay triangulation of class $\mathcal{Y}$ points. By construction, PE-PCDs allow us to work with Delaunay triangles one at a time. Moreover, the distribution of the edge density of the reflexivity and underlying graphs is shown to be geometry invariant for uniform data in one (Delaunay) triangle. This geometry invariance property allows us to work with the standard equilateral triangle, and utilize the inherent symmetry in such triangles, and thereby perform the calculations for the mean and variance of the asymptotic normal distribution as a function of the expansion parameter for once, and then conveniently extend the results to the more general case of multiple Delaunay triangles, i.e., the full support which is the convex hull of the $\mathcal{Y}$ points. We also compare the distributions of the edge density of the reflexivity and underlying graphs and that of the arc density of PE-PCDs.

We provide some preliminaries and the theoretical framework for random graphs and digraphs in Section 2, define PCDs and the associated reflexivity and underlying graphs and their edge densities (in a general framework and then specifically for the PE-PCDs) in Section 3, provide the asymptotic distribution of the edge density of the graphs associated with the PE-PCDs for uniform data in the one triangle case in Section 4. We treat the multiple triangle case in Section 5, provide the discussion and conclusions in Section 6, and the tedious calculations and long proofs are deferred to the Appendix and/or the technical report by Ceyhan [7].

2. Preliminaries

Let $\mathbb{Z}^+$ be the set of positive integers, $n \in \mathbb{Z}^+$, and $[n] := \{1, 2, \ldots, n\}$. Let also $\mathcal{G}_n$ (resp. $\boldsymbol{\mathcal{D}}_n$) denote the set of all simple graphs $G_n = (V, E)$ (resp. digraphs $\mathbf{D}_n = (V, A)$) with vertex set $V = [n]$ and edge set E (resp. arc set A). A simple graph is an undirected graph with no loops and no parallel edges. We often write the edge between i and j as the unordered pair ij for $i \neq j$ and say that i and j are adjacent. On the other hand, a simple digraph is a directed graph with no loops and no parallel arcs in the same direction. We write the arc from i to j as the ordered pair (i, j) for $i \neq j$ and say that j is adjacent to i.

Technically, a random graph (resp. digraph) is a probability space of the form $\mathbf{G}_n = (\mathcal{G}_n, \mathcal{F}, P)$ (resp. $\mathbf{D}_n = (\boldsymbol{\mathcal{D}}_n, \mathcal{F}, P)$) where $n \in \mathbb{Z}^+$ and P is the probability measure associated with $\mathcal{G}_n$ (resp. $\boldsymbol{\mathcal{D}}_n$). In fact, one should define the random graph (resp. digraph) as a graph-valued (resp. digraph-valued) random variable which is a measurable function from a probability space into $\mathcal{G}_n$ (resp. $\boldsymbol{\mathcal{D}}_n$). However, all our concern is the distribution of some invariant of such random variables which induces a probability measure on $\mathcal{G}_n$ (resp. $\boldsymbol{\mathcal{D}}_n$), hence the abuse of terminology. For random graphs and digraphs as the probability spaces defined above, the corresponding σ-algebra, $\mathcal{F}$, is the total σ-algebra; i.e., we assume that all subsets of the sample space $\mathcal{G}_n$ (resp. $\boldsymbol{\mathcal{D}}_n$) are measurable. We will drop $\mathcal{F}$ in the random graph and digraph notation when there is no concern of ambiguity.

The most popular and general random graphs are the Erdős–Rényi random graphs, $\mathbf{G}(n, p) = (\mathcal{G}_n, P)$ with

$$P(G) = p^{n_e}(1-p)^{\binom{n}{2}-n_e}, \quad G \in \mathcal{G}_n$$

where $p \in [0, 1]$ is the probability that an edge occurs between any two vertices, and $n_e = |E|$ (i.e., n_e is the number of edges in G). That is, the $\binom{n}{2}$ possible edges occur each with probability p and

independently of each other. Similarly the corresponding digraph can be defined as $\mathbf{D}(n, p) = (\mathcal{D}_n, P)$ with

$$P(D) = p_a^{n_a}(1 - p_a)^{n(n-1)-n_a}, \quad D \in \mathcal{D}_n$$

where $p_a \in [0, 1]$ is the probability that an arc occurs from a vertex to another, and $n_a = |A|$ (i.e., n_a is the number of arcs in D). Hence, the Erdős–Rényi type random digraph is obtained when each of $n(n-1)$ possible arcs occurs independently with probability p_a. To make a distinction, we will call $p = P(ij \in E, i \neq j)$ as the *edge probability* and $p_a = P((i, j) \in A, i \neq j)$ as the *arc probability*.

Beer et al. [2] provide a classification of isomorphism-invariant random graphs according to where the randomness resides. In particular, if only the edges occur randomly (but vertices are deterministic), then the graph is called an *edge random graph*; if only the vertices occur randomly (but edges are deterministic given the vertices), then the graph is a *vertex random graph*; and if both edges and vertices occur randomly, then the graph is the most general form called *vertex-edge random graph*. A random graph is isomorphism-invariant, if $P(G) = P(H)$ when G and H are in $\mathcal{G}_n$ and are isomorphic. The vertex-edge random graphs are defined as follows: Let $(\mathcal{V}, \mu)$ be a probability space, and $\psi : \mathcal{V} \times \mathcal{V} \to [0, 1]$ a symmetric function. Then the vertex-edge random graph $\mathbf{G}(n, \mathcal{V}, \mu, \psi)$ is the random graph $(\mathcal{G}_n, P)$ with

$$P(G) = \int P_{\mathbf{v}}(G)\mu(d\mathbf{v}) \tag{1}$$

where $P_{\mathbf{v}}$ is the probability measure associated with the edges conditional on vertices, and $\mu(d\mathbf{v})$ is the product integrator, $\mu(dv_1)\mu(dv_2)\ldots\mu(dv_n)$, on $\mathcal{V}^n$ (associated with vertices). That is, the vertex-edge random graph $\mathbf{G}(n, \mathcal{V}, \mu, \psi)$ is generated as follows: First n points are drawn iid from $\mathcal{V}$ with distribution μ, denoted as $\mathbf{V} = \{V_1, \ldots, V_n\}$. Then, conditional on $\mathbf{V}$, an edge is inserted between each pair of vertices i and j ($i \neq j$) independently with probability $\psi(V_i, V_j)$. If $\psi(x, y) \in \{0, 1\}\mu$-almost everywhere, then the vertex-edge random graph boils down to a vertex random graph. On the other hand, if $\mu(\mathbf{V}) = 1$ (i.e., μ is a trivial probability measure on $\mathcal{V}$), then the vertex-edge random graph reduces to an edge random graph. Hence, vertex random graphs and edge random graphs are special cases of the more encompassing family of vertex-edge random graphs.

The above classification can easily be extended to the (isomorphism-invariant) random digraphs by replacing edges with arcs and graphs with digraphs, but the function ψ would now be non-symmetric and is denoted as ψ_d (if it were symmetric, one does not have to resort to digraphs, simply the corresponding graph would be sufficient to work with).

A prominent example of vertex random graphs is the RGGs, which are extensively studied in literature [22]. For RGGs we have $\mathcal{V} = \mathbb{R}^k$: n points are generated iid from the probability distribution, μ, on $\mathbb{R}^k$, and an edge is inserted between two points (vertices) when they are within a certain distance t of each other. Thus $\psi(x, y) = \mathbf{I}(d(x, y) \leq t)$ where $d(\cdot, \cdot)$ is a distance in $\mathbb{R}^k$ and $\mathbf{I}(\cdot)$ is the indicator function.

The reflexivity and underlying graphs (which are the graphs of our concern in this article) of a digraph are the graphs obtained by replacing arcs with edges. If each arc $(i, j) \in A$ is replaced by an edge avoiding multi-edges, then we obtain the *underlying graph* of the digraph [11]. We also introduce a graph obtained by replacing each symmetric or reflexive arc, $\{(i, j), (j, i)\} \subset A$ by the edge ij and remove the arcs of the form $(i, j) \in A$ but $(j, i) \notin A$, and call the so-obtained graph *reflexivity graph* of the digraph. More specifically, the underlying graph for $\mathbf{D}_n = (V, A)$ is the graph $G_n^{\text{und}} = (V, E_{\text{und}})$ where E_{und} is the set of edges such that $ij \in E_{\text{und}}$ iff $(i, j) \in A$ ***or*** $(j, i) \in A$. The reflexivity graph of digraph $\mathbf{D}_n = (V, A)$ is the graph $G_n^{\text{ref}} = (V, E_{\text{ref}})$ where E_{ref} is the set of edges such that $ij \in E_{\text{ref}}$ iff $(i, j) \in A$ ***and*** $(j, i) \in A$. The sub- and super-scripting with "ref" and "und" are based on the corresponding terms of reflexivity and underlying graphs, respectively.

The main characteristic that makes the asymptotic study of the distribution of edge and arc densities related to the PCDs convenient is that these densities are U-statistics, whose asymptotic distribution is first studied by Hoeffding [15]. Being loyal to his notation (when there is no conflict with the ones in this article), in general, let $X_1, \ldots, X_n$ be a sequence of iid random variables and $\Phi(x_1, \ldots, x_k)$ be a function of $k(\leq n)$ variables. A statistic of the form

$$U_{n,k} = \sum \Phi(X_{i_1}, \ldots, X_{i_k})/n_{(k)} \tag{2}$$

where summation is over all permutations of the k distinct indices $i_1, \ldots, i_k$ and $n_{(k)} = n(n-1)\ldots(n-k+1)$ is called a U-statistic. Furthermore, if $X_1, \ldots, X_n$ are from the same distribution F, then $U_{n,k}$ is an unbiased estimate of the parameter $\theta_F = \int \ldots \int \Phi(X_1, \ldots, X_k) dF(x_1) \ldots dF(x_k)$ [15]. Hoeffding [15] shows that if $X_1, \ldots, X_n$ are from the same distribution, F and $\Phi(x_1, \ldots, x_k)$ is independent of the sample size n, then $\sqrt{n}(U_{n,k} - \theta)$ converges to a normal distribution provided that $\mathbf{E}[\Phi^2(X_1, \ldots, X_k)]$ exists. He also provides the asymptotic variance of the U-statistics [15] and a SLLN for them [16], and an exponential upper bound for the tail probabilities [17]. Letting $\Phi_0 := \frac{1}{k!}\sum_0 \Phi(X_{i_1}, \ldots, X_{i_k})$ where $\sum_0$ is taken over all permutations $(i_1, \ldots, i_k)$ of $(1, \ldots, k)$, $U_{n,k}$ can also be written as

$$U_{n,k} = \binom{n}{k}^{-1} \sum \Phi_0(X_{i_1}, \ldots, X_{i_k}) \tag{3}$$

where the summation is over $0 \le i_1 < i_2 < \cdots < i_k \le n$. Notice that the kernel $\Phi_0(x_1, \ldots, x_k)$ is symmetric function of k variables. We will mostly use the form in Eq. (3) in our analysis henceforth.

3. Edge density of underlying and reflexivity graphs based on PCDs

3.1. PCDs and the related graphs

Let $(\mathscr{V}, d)$ be a metric space equipped with a probability measure μ. Consider $N : \mathscr{V} \to \mathscr{P}(\mathscr{V})$, where $\mathscr{P}(\cdot)$ represents the power set. Then, given $\mathscr{Y}_m \subset \mathscr{V}$, the *proximity map* $N(\cdot)$ associates with each point $x \in \mathscr{V}$ a *proximity region* $N(x) \subseteq \mathscr{V}$. The region $N(x)$ is defined in terms of the distance between x and $\mathscr{Y}_m$. In what follows, we suppress the dependence on $\mathscr{Y}_m$ till Section 3.4 for notational convenience. Define the random PCD, $\mathbf{D}_n$, with vertex set $[n]$ and arc set A by $(i,j) \in A \iff X_j \in N(X_i)$ where point X_i is said to "catch" point X_j. Notice that we denote the vertices by $[n]$ or $\mathscr{X}_n$ for PCDs interchangeably, since it is a simple matter to connect $[n]$ to $\mathscr{X}_n$ by mapping $i \to X_i$ for $i = 1, 2, \ldots, n$. The random digraph $\mathbf{D}_n$ depends on the (joint) distribution of the X_i points and on the map $N(\cdot)$. That is, a PCD is a vertex random digraph $\mathbf{D}(n, \mathscr{V}, \mu, \psi_d)$ with $\psi_d(x, y) = \mathbf{I}(y \in N(x))$ and

$$P(D) = \int \mathbf{I}(\mathbf{D}(n, \mathscr{V}, \mu, \psi_d) = D)\mu(d\mathbf{v}), \quad D \in \mathscr{D}_n$$

where $\mu(d\mathbf{v})$ is as in Eq. (1). The adjective *proximity* – for the catch digraph $\mathbf{D}_n$ and for the map $N(\cdot)$ – comes from thinking of the region $N(x)$ as representing those points in $\mathscr{V}$ "close" to x [27,18]. The Γ_1-region $\Gamma_1(\cdot, N) : \mathscr{V} \to \mathscr{P}(\mathscr{V})$ associates the region $\Gamma_1(x, N) := \{z \in \mathscr{V} : x \in N(z)\}$ with each point $x \in \mathscr{V}$. A Γ_1-region is sort of a "dual" of the corresponding proximity region and is closely associated with domination number being equal to one [6]. Moreover, the PCD, $\mathbf{D}(n, \mathscr{V}, \mu, \psi_d)$, can be represented equivalently as $\mathbf{D}(n, \mathscr{V}, \mu, \widehat{\psi}_d)$ with $\widehat{\psi}_d(x, y) = \mathbf{I}(x \in \Gamma_1(y, N))$. Hence, $(i,j) \in A$ iff $X_i \in \Gamma_1(X_j, N)$. If $X_1, X_2, \ldots, X_n$ are $\mathscr{V}$-valued random variables, then the $N(X_i)$ (and $\Gamma_1(X_i, N)$), $i = 1, 2, \ldots, n$ are random sets. If the X_i are iid, then so are the functionals of the random sets $N(X_i)$ (and $\Gamma_1(X_i, N)$).

Consider the vertex random PCD, $\mathbf{D}_n$. The *reflexivity graph*, $\mathbf{G}_{\text{ref}}$, of $\mathbf{D}_n$ with the vertex set V and the edge set E_{ref} is defined by $ij \in E_{\text{ref}}$ iff $(i,j) \in A$ and $(j,i) \in A$. Likewise, the *underlying graph*, $\mathbf{G}_{\text{und}}$, of $\mathbf{D}_n$ with the vertex set V and the edge set E_{und} is defined by $ij \in E_{\text{und}} \iff (i,j) \in A$ or $(j,i) \in A$. Then $ij \in E_{\text{ref}}$ iff "$X_j \in N(X_i)$ and $X_i \in N(X_j)$" iff "$X_j \in N(X_i)$ and $X_j \in \Gamma_1(X_i N)$" iff $X_j \in N(X_i) \cap \Gamma_1(X_i, N)$. Similarly, $ij \in E_{\text{und}}$ iff $X_j \in N(X_i) \cup \Gamma_1(X_i, N)$. Hence the reflexivity and underlying graphs are generalized versions of random geometric graphs with the metric and the threshold distance are implicit in the defining regions $N(\cdot)$ (and $\Gamma_1(\cdot, N)$). Furthermore, these graphs are also vertex random graphs. In the vertex random graph notation of Section 2, we have $\mathbf{G}_{\text{ref}}(n, \mathscr{V}, \mu, \psi_{\text{ref}})$ with $\psi_{\text{ref}}(x, y) = \mathbf{I}(y \in N(x) \cap \Gamma_1(x, N))$ for the reflexivity graph; and we have $\mathbf{G}_{\text{und}}(n, \mathscr{V}, \mu, \psi_{\text{und}})$ with $\psi_{\text{und}}(x, y) = \mathbf{I}(y \in N(x) \cup \Gamma_1(x, N))$ for the underlying graph.

Proposition 3.1. *The* PCD*s and the associated reflexivity and underlying graphs are isomorphism-invariant.*

Proof. Suppose D and D' are two isomorphic PCDs. Since the n vertices are drawn iid from $(\mathscr{V}, P)$, we would have $P((i,j) \in A(D)) = P((i',j') \in A(D'))$ for any $i \neq j$ and $i' \neq j'$. Therefore, $P(D) = P(D')$. Similarly, the corresponding reflexivity (and underlying) graphs for D and D' would also be isomorphic, hence would have the same probability, hence are isomorphism-invariant. ■

3.2. Arc density of the PCDs

The arc density of the PCD, $\mathbf{D}_n = (V, A) \in (\mathscr{D}_n, P)$, is denoted as $\rho(\mathbf{D}_n)$. Recall that $\psi_d(x,y) = \mathbf{I}(y \in N(x))$. Then the arc density $\rho(\mathbf{D}_n)$ can be written as

$$\rho(\mathbf{D}_n) = \frac{|A|}{n(n-1)} = \frac{1}{n\,(n-1)} \sum\sum_{i \neq j} \psi_d(X_i, X_j) = \frac{2}{n\,(n-1)} \sum\sum_{i<j} h(X_i, X_j) \tag{4}$$

where $2\,h(X_i, X_j) = \psi_d(X_i, X_j) + \psi_d(X_j, X_i) = \mathbf{I}(X_j \in N(X_i)) + \mathbf{I}(X_i \in N(X_j))$ is the number of arcs between vertices X_i and X_j in $\mathbf{D}_n$. For $X_i \overset{iid}{\sim} F$, $i = 1, 2, \ldots, n$, with support of F being in $\mathscr{V}$, $h(x,y)$ is a symmetric function of two variables. Moreover, $\rho(\mathbf{D}_n)$ is a random variable $\rho : \mathscr{D}_n \to [0,1]$ and depends on n, F, and $N(\cdot)$. But $\mathbf{E}[\rho(\mathbf{D}_n)]$ only depends on F and $N(\cdot)$. That is,

$$0 \leq \mathbf{E}[\rho(\mathbf{D}_n)] = \frac{2}{n\,(n-1)} \sum\sum_{i<j} \mathbf{E}[h(X_i, X_j)] = \mathbf{E}[h(X_1, X_2)] \tag{5}$$

where $2\,\mathbf{E}[h(X_1, X_2)] = \mathbf{E}[\psi_d(X_1, X_2) + \psi_d(X_2, X_1)] = 2\,P(X_2 \in N(X_1)) = 2\,p_a$. Hence $\mathbf{E}[\rho(\mathbf{D}_n)] = \mathbf{E}[h(X_1, X_2)] = p_a$, which is the *arc probability* for the PCD, $\mathbf{D}_n$, since $p_a = P(X_j \in N(X_i)) = P((i,j) \in A(\mathbf{D}_n))$ for $i \neq j$. Furthermore,

$$0 \leq \mathbf{Var}[\rho(\mathbf{D}_n)] = \frac{4}{n^2\,(n-1)^2} \mathbf{Var}\left[\sum\sum_{i<j} h(X_i, X_j)\right]. \tag{6}$$

Expanding this expression, we have

$$\mathbf{Var}[\rho(\mathbf{D}_n)] = \frac{2}{n\,(n-1)} \mathbf{Var}[h(X_1, X_2)] + \frac{4(n-2)}{n\,(n-1)} \mathbf{Cov}[h(X_1, X_2), h(X_1, X_3)]$$

with

$$\mathbf{Var}[h(X_1, X_2)] = (p_a - p_a^2) = p_a\,(1 - p_a)$$

and

$$\begin{aligned}\mathbf{Cov}[h(X_1, X_2), h(X_1, X_3)] &= \mathbf{E}[h(X_1, X_2) h(X_1, X_3)] \\ &\quad - \mathbf{E}[h(X_1, X_2)]\,\mathbf{E}[h(X_1, X_3)] = \mathbf{E}[h(X_1, X_2) h(X_1, X_3)] - p_a^2,\end{aligned}$$

where

$$\begin{aligned}&4\,\mathbf{E}[h(X_1, X_2) h(X_1, X_3)] \\ &\quad = \mathbf{E}[(\psi_d(X_1, X_2) + \psi_d(X_2, X_1))(\psi_d(X_1, X_3) + \psi_d(X_3, X_1))] \\ &\quad = \mathbf{E}[\mathbf{I}(X_2 \in N(X_1))\mathbf{I}(X_3 \in N(X_1)) + \mathbf{I}(X_2 \in N(X_1))\mathbf{I}(X_1 \in N(X_3)) \\ &\qquad + \mathbf{I}(X_1 \in N(X_2)\mathbf{I}(X_3 \in N(X_1))] + \mathbf{I}(X_1 \in N(X_2))\mathbf{I}(X_1 \in N(X_3))] \\ &\quad = \mathbf{E}[\mathbf{I}(\{X_2, X_3\} \subset N(X_1)) + \mathbf{I}(X_2 \in N(X_1))\mathbf{I}(X_3 \in \Gamma_1(X_1, N)) \\ &\qquad + \mathbf{I}(X_2 \in \Gamma_1(X_1)\mathbf{I}(X_3 \in N(X_1))] + \mathbf{I}(X_2 \in \Gamma_1(X_1, N))\mathbf{I}(X_3 \in \Gamma_1(X_1, N))] \\ &\quad = P(\{X_2, X_3\} \subset N(X_1)) + 2\,P(X_2 \in N(X_1), X_3 \in \Gamma_1(X_1, N)) \\ &\qquad + P(\{X_2, X_3\} \subset \Gamma_1(X_1, N)).\end{aligned}$$

The arc density of the PCDs converges weakly to a normal distribution as shown in the next theorem.

Theorem 3.2. *Let* $X_i \overset{iid}{\sim} F$ *and* $\mathbf{D}_n$ *be the* PCD *which is based on the proximity map* $N(\cdot)$ *and has vertex set* $\mathcal{X}_n$. *The arc density,* $\rho(\mathbf{D}_n)$, *of the* PCD, $\mathbf{D}_n$, *defined in Eq. (4) is a one-sample U-statistic of degree* 2 *and is an unbiased estimator of* p_a. *Additionally, if* $\nu_a = \mathbf{Cov}\left[h(X_i, X_j), h(X_i, X_k)\right] > 0$ *for all* $i \neq j \neq k, i, j, k = 1, 2, \ldots, n$, *then* $\sqrt{n}\left(\rho(\mathbf{D}_n) - p_a\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_a)$ *as* $n \to \infty$ *where* $\xrightarrow{\mathcal{L}}$ *stands for convergence in law or distribution and* $\mathcal{N}(\mu, \sigma^2)$ *stands for the normal distribution with mean* μ *and variance* σ^2.

Proof. The arc probability, p_a, for the random digraph $\mathbf{D}_n$ is an estimable parameter of degree 2, since $\mathbf{E}[h(X_1, X_2)] = p_a$ (so $h(X_1, X_2)$ is unbiased for p_a) and $h(x, y)$ is a symmetric function of two variables. As seen in Eq. (4), $\rho(\mathbf{D}_n)$ is of the form $\binom{n}{2}^{-1} \sum h(X_i, X_j)$ where the summation is over all integers (i, j) with $i < j$ from $(1, 2, \ldots, n)$. Hence $\rho(\mathbf{D}_n)$ is a U-statistic with the symmetric kernel $h(x, y)$, and is an unbiased estimator of p_a (by Eq. (5)). Note also that X_i are iid from F for $i = 1, \ldots, n$, and $\psi_d(x, y)$ is independent of n. Moreover, $\mathbf{E}[\psi_d^2(X_1, X_2)] = \mathbf{E}[\mathbf{I}(X_2 \in N(X_1))] = P(X_2 \in N(X_1)) = p_a \leq 1$ exists, then by Theorem 7.1 of Hoeffding [15], it follows that $\sqrt{n}\left(\rho(\mathbf{D}_n) - p_a\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_a)$ provided that $\nu_a > 0$ where $4\,\nu_a = \mathbf{Cov}(h(X_1, X_2), h(X_1, X_3))$. ■

In the above theorem, $\nu_a > 0$ iff $P(\{X_2, X_3\} \subset N(X_1)) + 2\,P(X_2 \in N(X_1)), \mathbf{I}(X_3 \in \Gamma_1(X_3, N)) + P(\{X_2, X_3\} \subset \Gamma_1(X_1, N)) > 4\,p_a^2$. We can also obtain the joint distribution of $(h(X_1, X_2), h(X_1, X_3))$ (see Appendix 1).

Remark 3.3. Notice that for PCDs, the set of (place holders of) vertices $V = \mathcal{X}_n$ is a random sample from a distribution F (i.e., the vertices directly result from a random process), and the arcs are defined based on the random sets (i.e., proximity regions) $N(X_i)$ as described in Section 3.1. Hence the set of arcs A (indirectly) results from a random process such that $\psi_d(X_i, X_j)$ are identically distributed and $\psi_d(X_i, X_j)$ and $\psi_d(X_k, X_l)$ are independent for distinct i, j, k, l. We have $\mathbf{Cov}\left[h(X_i, X_j), h(X_i, X_k)\right] < \infty$, since both $\mathbf{E}\left[h(X_1, X_2)h(X_1, X_3)\right]$ and p_a are finite. Furthermore, asymptotic distribution of $\rho(\mathbf{D}_n)$ is non-degenerate whenever $\nu_a > 0$ and is degenerate with $\rho(\mathbf{D}_n) = p_a$ a.s. whenever $\nu_a = 0$. □

Remark 3.4. Showing $\rho(\mathbf{D}_n)$ is a U-statistic of degree 2 entails more than just asymptotic normality. In particular:

(i) By Theorem 5.1 of Hoeffding [15], we have $\mathbf{Cov}[h(X_1, X_2), h(X_1, X_3)] \leq \mathbf{Var}[h(X_1, X_2)]$.
(ii) By Theorem 5.2 of Hoeffding [15], we have

$$\frac{4}{n}\mathbf{Cov}[h(X_1, X_2), h(X_1, X_3)] \leq \mathbf{Var}[\rho(\mathbf{D}_n)] \leq \frac{2}{n}\mathbf{Var}[h(X_1, X_2)],$$

and $\mathbf{Var}[\rho(\mathbf{D}_n)]$ is a decreasing function of n as n increases, and if $\nu_a > 0$, $\mathbf{Var}[\rho(\mathbf{D}_n)]$ is of order $1/n$.
(iii) If $\nu_a > 0$, SLLN follows, since $\mathbf{E}[\psi_d(X_1, X_2)] = p_a$ is finite [16]; that is, $\rho(\mathbf{D}_n) \xrightarrow{a.s.} p_a$ as $n \to \infty$.
(iv) Finally, since $0 \leq \psi_d(X_1, X_2) \leq 1$, by Hoeffding [17], we have $P(\rho(\mathbf{D}_n) - p_a > t) \leq \exp(-2kt^2)$ where $k = \lfloor n/2 \rfloor$, the largest integer less than or equal to $n/2$. □

3.3. Edge density of the graphs based on PCDs

Let G_n^{ref} and G_n^{und} be the reflexivity and underlying graphs associated with the PCD, $\mathbf{D}_n$, respectively. Also let ρ_n^{ref} and ρ_n^{und} be the corresponding edge densities of the graphs, G_n^{ref} and G_n^{und}, respectively. One can express the edge density of G_n^{ref} as

$$\rho_n^{\text{ref}} = \frac{1}{n\,(n-1)} \sum \sum_{i \neq j} \psi_{\text{ref}}(X_i, X_j) = \frac{2}{n\,(n-1)} \sum \sum_{i<j} \psi_{\text{ref}}(X_i, X_j) \tag{7}$$

where $\psi_{\text{ref}}(X_i, X_j) = \mathbf{I}(X_j \in N(X_i) \cap \Gamma_1(X_i, N))$ is the indicator for the existence of an edge between X_i and X_j in $\mathbf{G}_n^{\text{ref}}$ or number of symmetric or reflexive arcs between X_i and X_j in $\mathbf{D}_n$. Similarly, the edge

density of $\mathbf{G}_n^{\text{und}}$ is

$$\rho_n^{\text{und}} = \frac{1}{n\,(n-1)} \sum\sum_{i\neq j} \psi_{\text{und}}(X_i, X_j) = \frac{2}{n\,(n-1)} \sum\sum_{i<j} \psi_{\text{und}}(X_i, X_j) \tag{8}$$

where $\psi_{\text{und}}(X_i, X_j) = \mathbf{I}(X_j \in N(X_i) \cup \Gamma_1(X_i, N))$ is the indicator for the existence of an edge between X_i and X_j in $\mathbf{G}_n^{\text{und}}$.

For $X_i \overset{iid}{\sim} F$, $i = 1, 2, \ldots, n$, $\psi_{\text{ref}}(x, y)$ (resp. $\psi_{\text{und}}(x, y)$) is a symmetric function of two variables. Moreover, ρ_n^{ref} and ρ_n^{und} are random variables that depend on n, F, and $N(\cdot)$. But $\mathbf{E}\left[\rho_n^{\text{ref}}\right]$ and $\mathbf{E}\left[\rho_n^{\text{und}}\right]$ only depend on F and $N(\cdot)$. That is, we have

$$0 \leq \mathbf{E}\left[\rho_n^{\text{ref}}\right] = \frac{2}{n\,(n-1)} \sum\sum_{i<j} \mathbf{E}[\psi_{\text{ref}}(X_i, X_j)] = \mathbf{E}\left[\psi_{\text{ref}}(X_1, X_2)\right] = p_{\text{ref}}. \tag{9}$$

Similarly,

$$0 \leq \mathbf{E}\left[\rho_n^{\text{und}}\right] = \frac{2}{n\,(n-1)} \sum\sum_{i<j} \mathbf{E}[\psi_{\text{und}}(X_i, X_j)] = \mathbf{E}\left[\psi_{\text{und}}(X_1, X_2)\right] = p_{\text{und}}. \tag{10}$$

Notice that $p_{\text{ref}} = P(X_j \in N(X_i) \cap \Gamma_1(X_i, N))$ for $i \neq j$ is edge probability for the reflexivity graph $\mathbf{G}_n^{\text{ref}}$, and it is also the symmetric or reflexive arc probability for the PCD, $\mathbf{D}_n$. Likewise, $p_{\text{und}} = P(X_j \in N(X_i) \cup \Gamma_1(X_i, N))$ for $i \neq j$ is the edge probability for the underlying graph $\mathbf{G}_n^{\text{und}}$.

Moreover,

$$0 \leq \mathbf{Var}\left[\rho_n^{\text{ref}}\right] = \frac{4}{n^2\,(n-1)^2} \mathbf{Var}\left[\sum\sum_{i<j} \psi_{\text{ref}}(X_i, X_j)\right]. \tag{11}$$

Expanding this expression, we get

$$\mathbf{Var}\left[\rho_n^{\text{ref}}\right] = \frac{2}{n\,(n-1)} \mathbf{Var}\left[\psi_{\text{ref}}(X_1, X_2)\right] + \frac{4\,(n-2)}{n\,(n-1)} \mathbf{Cov}\left[\psi_{\text{ref}}(X_1, X_2),\, \psi_{\text{ref}}(X_1, X_3)\right]$$

with

$$\mathbf{Var}\left[\psi_{\text{ref}}(X_1, X_2)\right] = p_{\text{ref}} - p_{\text{ref}}^2 = p_{\text{ref}}\,(1 - p_{\text{ref}})$$

and

$$\mathbf{Cov}\left[\psi_{\text{ref}}(X_1, X_2),\, \psi_{\text{ref}}(X_1, X_3)\right] = \mathbf{E}\left[\psi_{\text{ref}}(X_1, X_2)\psi_{\text{ref}}(X_1, X_3)\right] - \mathbf{E}\left[\psi_{\text{ref}}(X_1, X_2)\right]\mathbf{E}\left[\psi_{\text{ref}}(X_1, X_3)\right].$$

Since $\mathbf{E}\left[\psi_{\text{ref}}(X_1, X_2)\right] = \mathbf{E}\left[\psi_{\text{ref}}(X_1, X_3)\right] = p_{\text{ref}}$ and,

$$\begin{aligned}
\mathbf{E}\left[\psi_{\text{ref}}(X_1, X_2)\psi_{\text{ref}}(X_1, X_3)\right] &= \mathbf{E}[\mathbf{I}(X_2 \in N(X_1) \cap \Gamma_1(X_1, N))\,.\,\mathbf{I}(X_3 \in N(X_1) \cap \Gamma_1(X_1, N))] \\
&= P(X_2 \in N(X_1) \cap \Gamma_1(X_1, N)\,,\, X_3 \in N(X_1) \cap \Gamma_1(X_1, N)) \\
&= P(\{X_2, X_3\} \subset N(X_1) \cap \Gamma_1(X_1, N)),
\end{aligned}$$

it follows that

$$\mathbf{Cov}\left[\psi_{\text{ref}}(X_1, X_2),\, \psi_{\text{ref}}(X_1, X_3)\right] = P(\{X_2, X_3\} \subset N(X_1) \cap \Gamma_1(X_1, N)) - p_{\text{ref}}^2.$$

Similarly, we have,

$$0 \leq \mathbf{Var}\left[\rho_n^{\text{und}}\right] = \frac{2}{n\,(n-1)} \mathbf{Var}\left[\psi_{\text{und}}(X_1, X_2)\right] + \frac{4\,(n-2)}{n\,(n-1)} \mathbf{Cov}[\psi_{\text{und}}(X_1, X_2),\, \psi_{\text{und}}(X_1, X_3)] \tag{12}$$

with

$$\mathbf{Var}\left[\psi_{\text{und}}(X_1, X_2)\right] = p_{\text{und}} - p_{\text{und}}^2 = p_{\text{und}}\,(1 - p_{\text{und}})$$

and

$$\mathbf{Cov}\left[\psi_{\text{und}}(X_1, X_2),\, \psi_{\text{und}}(X_1, X_3)\right] = P(\{X_2, X_3\} \subset N(X_1) \cup \Gamma_1(X_1, N)) - p_{\text{und}}^2.$$

The edge density of the reflexivity and underlying graphs also converge in distribution to normality as proved below.

Theorem 3.5. *Let $X_i \overset{iid}{\sim} F$ and $\mathbf{D}_n$ be the* PCD *which is based on the proximity map $N(\cdot)$ and has vertex set $\mathcal{X}_n$. The edge density, ρ_n^{ref}, of the reflexivity graph of the* PCD *$\mathbf{D}_n$, $\mathbf{G}_n^{ref}$, is a one-sample U-statistic of degree 2 and is an unbiased estimator of p_{ref}. Additionally, if $\nu_{ref} := \mathbf{Cov}\left[\psi_{ref}(X_i, X_j), \psi_{ref}(X_i, X_k)\right] > 0$ for all $i \neq j \neq k, i, j, k = 1, 2, \ldots, n$, then $\sqrt{n}\,(\rho_n^{ref} - p_{ref}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_{ref})$ as $n \to \infty$. The same conclusions hold for the underlying graph with 'ref's being replaced with 'und's.*

Proof. For the reflexivity graph, $\mathbf{G}_n^{\text{ref}}$ of the PCD $\mathbf{D}_n$, the edge probability is an estimable parameter of degree 2, since $\mathbf{E}[\psi_{\text{ref}}(X_1, X_2)] = p_{\text{ref}}$ (so $\psi_{\text{ref}}(X_1, X_2)$ is unbiased for p_{ref}) and $\psi_{\text{ref}}(x, y)$ is a symmetric function of two variables. As seen in Eq. (7), ρ_n^{ref} is a U-statistic with the symmetric kernel $\psi_{\text{ref}}(x, y)$, and is an unbiased estimator of p_{ref} (by Eq. (9)). Note also that X_i are iid from F for $i = 1, \ldots, n$, and $\psi_{\text{ref}}(x, y)$ is independent of n. Moreover, $\mathbf{E}[\psi_{\text{ref}}^2(X_1, X_2)] = \mathbf{E}[\mathbf{I}(X_2 \in N(X_1) \cap \Gamma_1(X_1, N))] = P(X_2 \in N(X_1) \cap \Gamma_1(X_1, N)) = p_{\text{ref}} \leq 1$ exists, then by Theorem 7.1 of Hoeffding [15], it follows that $\sqrt{n}\,(\rho_n^{\text{ref}} - p_{\text{ref}}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_{\text{ref}})$ provided that $\nu_{\text{ref}} > 0$ where $4\,\nu_{\text{ref}} = \mathbf{Cov}[\psi_{\text{ref}}(X_1, X_2), \psi_{\text{ref}}(X_1, X_3)]$. The same arguments hold for the underlying graph, with 'ref's replaced with 'und's and $\cap$'s replaced with $\cup$'s. ■

In the above theorem, $\nu_{\text{ref}} > 0$ iff $P(\{X_2, X_3\} \subset N(X_1) \cap \Gamma_1(X_1, N)) > p_{\text{ref}}^2$. Similarly, $\nu_{\text{und}} > 0$ iff $P(\{X_2, X_3\} \subset N(X_1) \cup \Gamma_1(X_1, N)) > p_{\text{und}}^2$. The comments in Remark 3.3 and results in Remark 3.4 for the arc density of the PE-PCDs hold for the edge density of the reflexivity and underlying graphs of the PE-PCDs as well as with straightforward modifications.

Remark 3.6. In Theorem 3.2, we have $\mathbf{E}\left[|h(X_i, X_j)|^3\right] < \infty$, since $\mathbf{E}\left[|h(X_i, X_j)|^3\right] \leq 1$. Then for $\nu_a > 0$, letting $\Phi(t)$ be the cumulative distribution function for standard normal distribution, the sharpest rate of convergence of $\rho(\mathbf{D}_n)$ to the asymptotic normality is

$$\sup_{t \in \mathbb{R}} \left| P\left(\frac{\sqrt{n}(\rho(\mathbf{D}_n) - p_a)}{\sqrt{4\,\nu_a}} \leq t \right) - \Phi(t) \right| \leq 8\,K\,p_a\,(4\,\nu)^{-3/2}\,n^{-1/2} = K\,\frac{p_a}{\sqrt{n\,\nu_a^3}} \tag{13}$$

where K is a constant [4]. The sharpest rate for the reflexivity (resp. underlying) graphs in Theorem 4.2 can be obtained by replacing p_a with p_{ref} (resp. p_{und}) and ν_a with ν_{ref} (resp. ν_{und}). □

The joint distributions of $(\psi_{\text{ref}}(X_1, X_2), \psi_{\text{ref}}(X_1, X_3))$ and $(\psi_{\text{und}}(X_1, X_2), \psi_{\text{und}}(X_1, X_3))$ are also available (see Appendix 1 for details).

Remark 3.7. Note that $2\,h(X_i, X_j) = \psi_{\text{ref}}(X_i, X_j) + \psi_{\text{und}}(X_i, X_j)$, since if $\psi(X_i, X_j) = \psi(X_j, X_i) = 0$, then $2\,h(X_i, X_j) = 0$, and $\psi_{\text{ref}}(X_i, X_j) = \psi_{\text{und}}(X_i, X_j) = 0$; if $\psi(X_i, X_j) = \psi(X_j, X_i) = 1$, then $2\,h(X_i, X_j) = 2$, and $\psi_{\text{ref}}(X_i, X_j) = \psi_{\text{und}}(X_i, X_j) = 1$; and if $\psi(X_i, X_j) = 0$ and $\psi(X_j, X_i) = 1$, then $2\,h(X_i, X_j) = 1$, and $\psi_{\text{ref}}(X_i, X_j) = 0$ and $\psi_{\text{und}}(X_i, X_j) = 1$; by symmetry, the same holds when $\psi(X_i, X_j) = 1$ and $\psi(X_j, X_i) = 0$.

Furthermore, based on Proposition 3.1, each PCD of order n with the same number of arcs (i.e., with equal arc densities) has the same probability of occurring. The same holds for the underlying (and reflexivity) graphs with the same number of edges. □

3.4. Proportional-edge proximity maps and the associated regions

Now we introduce the PCD family for which we explicitly specify the form of the proximity region $N(\cdot)$. Let $\mathscr{V} = \mathbb{R}^2$ and $\mathcal{Y}_m = \{y_1, y_2, \ldots, y_m\} \subset \mathbb{R}^2$ be m points in general position such that no more than three points are co-circular (i.e., lie on the same circle). Then the Delaunay triangulation based on $\mathcal{Y}_m$ is unique and the Delaunay triangles partition the convex hull of $\mathcal{Y}_m$. See Fig. 1 for 10 $\mathcal{Y}$ iid points uniformly generated in the unit square and the corresponding Delaunay triangulation which consists of 13 triangles. We will restrict our attention to iid uniform $\mathcal{X}$ points in the convex hull of $\mathcal{Y}$ points. See Fig. 2 for the 200 $\mathcal{X}$ points iid generated in the unit square where 77 of the $\mathcal{X}$ points lie inside the convex hull of $\mathcal{Y}$ points in Fig. 1.

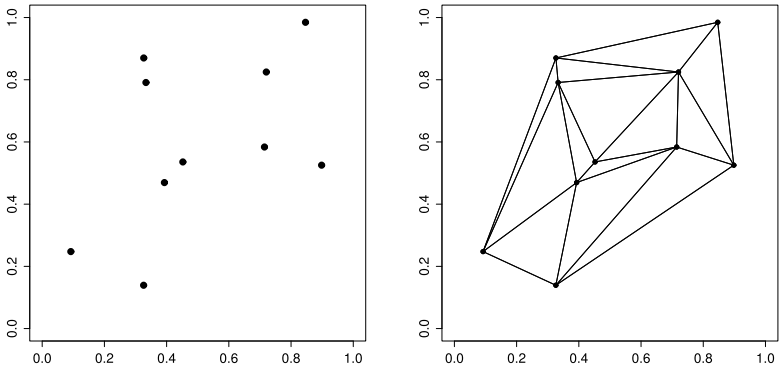

Fig. 1. A realization of 10 $\mathcal{Y}$ points generated iid uniformly on the unit square (left) and the corresponding Delaunay triangulation which consists of 13 triangles.

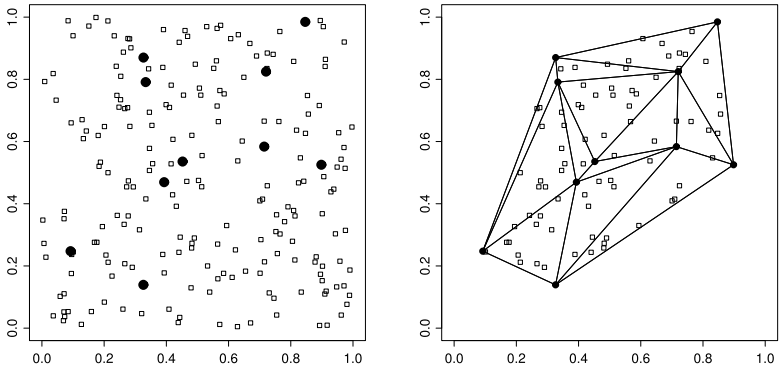

Fig. 2. A realization of 200 $\mathcal{X}$ points (squares) generated iid uniformly on the unit square (left) and the $\mathcal{X}$ points restricted to the convex hull of $\mathcal{Y}$ points (right) where the 10 $\mathcal{Y}$ points (filled circles) are from Fig. 1.

For simplicity, we first consider the $m = 3$ case, i.e., we let $\mathcal{Y}_3 = \{y_1, y_2, y_3\} \subset \mathbb{R}^2$ be three non-collinear points. Then Delaunay triangulation of $\mathcal{Y}$ points yields a single triangle. Denote this triangle (including the interior) by $T(\mathcal{Y}_3)$. For $r \in [1, \infty]$ define $N_{PE}(x, r)$ to be the *proportional-edge* proximity map with parameter r and $\Gamma_1(x, r) := \Gamma_1(x, N_{PE}(\cdot, r))$ to be the corresponding Γ_1-region as follows; see also Figs. 3 and 4. Let "vertex regions" $R(y_1), R(y_2), R(y_3)$ partition $T(\mathcal{Y}_3)$ using segments from the center of mass of $T(\mathcal{Y}_3)$ to the edge midpoints. 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; $x \in R(v(x))$. If x falls on the boundary of two vertex regions, or at the center of mass, 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 (perpendicular) distance from $v(x)$ to $\ell(v(x), x)$. For $r \in [1, \infty)$ let $\ell_r(v(x), x)$ be the line parallel to $e(x)$ such that

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

Let $T_r(x)$ be the triangle similar to and with the same orientation as $T(\mathcal{Y}_3)$ having $v(x)$ as a vertex and $\ell_r(v(x), x)$ as the opposite edge. Then the proportional-edge proximity region $N_{PE}(x, r)$ is defined to be $T_r(x) \cap T(\mathcal{Y}_3)$. Notice that $N_{PE}(x, r)$ is parameterized by an expansion parameter r and is implicitly depending on the center of mass of the triangle.

Furthermore, let $\xi_i(x)$ be the line such that $\xi_i(x) \cap T(\mathcal{Y}_3) \neq \emptyset$ and $r\, d(y_i, \xi_i(x)) = d(y_i, \ell(y_i, x))$ for $i = 1, 2, 3$. Then $\Gamma_1(x, r) \cap R(y_i) = \{z \in R(y_i) : d(y_i, \ell(y_i, z)) \geq d(y_i, \xi_i(x))\}$, for $i = 1, 2, 3$. Hence $\Gamma_1(x, r) = \bigcup_{i=1}^{3} (\Gamma_1(x, r) \cap R(y_i))$. Notice that $r \geq 1$ implies $x \in N_{PE}(x, r)$ and $x \in \Gamma_1(x, r)$. Furthermore, $\lim_{r \to \infty} N_{PE}(x, r) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, and so we define $N_{PE}(x, \infty) = T(\mathcal{Y}_3)$

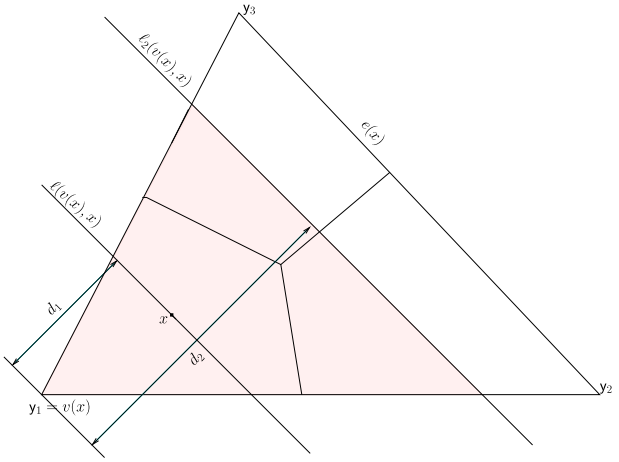

Fig. 3. Construction of proportional-edge proximity region, $N_{PE}(x, r = 2)$ (shaded region) for an $x \in R(\mathsf{y}_1)$ where $d_1 = d(v(x), \ell(v(x), x))$ and $d_2 = d(v(x), \ell_2(v(x), x)) = 2\, d(v(x), \ell(v(x), x))$.

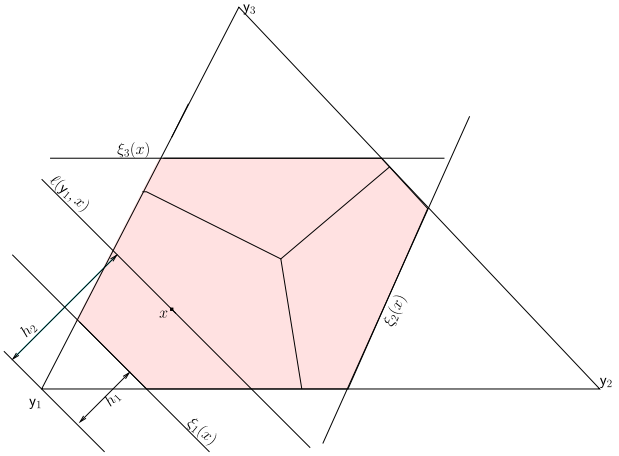

Fig. 4. Construction of the Γ_1-region, $\Gamma_1(x, r = 2)$ (shaded region) for an $x \in R(\mathsf{y}_1)$ where $h_1 = d(\mathsf{y}_1, \xi_1(x))$ and $h_2 = d(\mathsf{y}_1, \ell(\mathsf{y}_1, x)) = r\, h_1$.

for all such x. For $x \in \mathcal{Y}_3$, we define $N_{PE}(x, r) = \{x\}$ for all $r \in [1, \infty]$. Then, for $x \in R(\mathsf{y}_i)$, $\lim_{r\to\infty} \Gamma_1(x, r) = T(\mathcal{Y}_3) \setminus \{\mathsf{y}_j, \mathsf{y}_k\}$ for distinct i, j, and k.

Notice that $X_i \overset{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional probability density function f exists with support in $T(\mathcal{Y}_3)$, implies that the special cases in the construction of $N_{PE}(\cdot, r)$ – X falls on the boundary of two vertex regions, or at the center of mass, or $X \in \mathcal{Y}_3$ – occur with probability zero. Note that for such an F, $N_{PE}(x, r)$ is a triangle a.s. and $\Gamma_1(x, r)$ is a convex or nonconvex polygon.

PE-PCDs can be viewed as a digraph version of (generalized) random geometric graphs. The PE-PCD is based on a locally defined and restricted dissimilarity measure. In particular, the dissimilarity $d_m(\cdot, \cdot)$ between two points x and y is only defined if both x and y are in the same triangle. More specifically, $d_m(x, y) := |d(v(x), \ell(v(x), x)) - d(v(x), \ell(v(x), y))| = d(\ell(v(x), x), \ell(v(x), y))$. But notice that $d_m(y, x) = d(\ell(v(y), y), \ell(v(y), x))$ which may not equal $d_m(x, y)$. Indeed, $d_m(x, y) = d_m(y, x)$ iff $v(x) = v(y)\lambda$-almost everywhere where λ is the Lebesgue measure; that is,

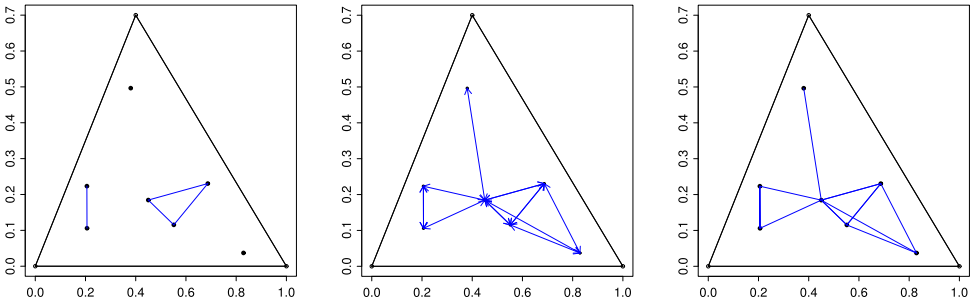

Fig. 5. Presented in the middle are the arcs for the PE-PCD with $r = 3/2$ for 7 $\mathcal{X}$ points iid uniformly generated in a triangle based on 3 $\mathcal{Y}$ points. Plotted in the left are the edges for the associated reflexivity graph and right are for the underlying graph.

$d_m(x, y) = d_m(y, x)$ when $v(x) \neq v(y)$ holds on a measure zero set in $T(\mathcal{Y}_3)$. Hence symmetry fails for $d_m(\cdot, \cdot)$. Furthermore, $d_m(x, x) = 0$ but $d_m(x, y)$ could be zero for $x \neq y$ (e.g., this would hold if $y \in \ell(v(x), x)$). Hence $d_m(x, y) = 0$ does not necessarily imply $x = y$ (although $d_m(x, y) = 0$ for $x \neq y$ holds on a measure zero set), hence coincidence axiom fails for $d_m(\cdot, \cdot)$ as well. However, non-negativity holds for $d_m(\cdot, \cdot)$, since for all $x, y \in T(\mathcal{Y}_3)$, $d_m(x, y) \geq 0$ and $d_m(y, x) \geq 0$. Finally triangle inequality may also fail for d_m, since $d_m(x, z) \geq d_m(x, y) + d_m(y, z)$ may hold when $v(x) \neq v(y)$. Here $d_m(x, y) = d(\ell(v(x), x), \ell(v(x), y))$ and $d_m(y, z) = d(\ell(v(y), y), \ell(v(y), z))$. Hence $d_m(x, y) + d_m(y, z) = d(\ell(v(x), x), \ell(v(x), y)) + d(\ell(v(y), y), \ell(v(y), z))$ which can be less than $d(\ell(v(x), x), \ell(v(x), z)) = d_m(x, z)$. However, d_m locally satisfies more of the distance axioms. In particular, if $x, y \in R(\mathrm{y}_i)$ (i.e., $v(x) = v(y)$), then $d_m(x, y) = d_m(y, x)$, hence symmetry holds. If $x, y, z \in R(\mathrm{y}_i)$ then the triangle inequality holds with $d_m(x, z) \leq d_m(x, y) + d_m(y, z)$. The threshold for this extension of random geometric graphs would be $d(v(x), \ell_r(v(x), x)) = rd(v(x), \ell(v(x), x))$ for $x \in R(v(x))$ if $d(v(x), \ell(v(x), y)) \geq d(v(x), \ell(v(x), x))$ and $d(v(x), \ell(v(x), x))$ otherwise. That is,

$$(x, y) \in A \text{ iff } \begin{cases} d_m(x, y) \leq d(v(x), \ell_r(v(x), x)) & \text{if } d(v(x), \ell(v(x), y)) \geq d(v(x), \ell(v(x), x)), \\ d_m(x, y) \leq d(v(x), \ell(v(x), x)) & \text{otherwise.} \end{cases} \tag{14}$$

The reflexivity and underlying graphs of PE-PCDs can also be viewed as generalizations of random geometric graphs with the dissimilarity measure $d_m(\cdot, \cdot)$. For example, in the reflexivity graph, we insert an edge between x and y iff $\{(x, y), (y, x)\} \subset A$. That is, $d_m(x, y)$ and $d_m(y, x)$ satisfy the threshold condition in Eq. (14).

3.5. Edge density of the graphs based on PE-PCDs

Let $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ be a random sample from a distribution F with support in $T(\mathcal{Y}_3)$. Let $\mathbf{D}_n(r)$ be the PE-PCD with vertex set $V = \mathcal{X}_n$ and arc set A defined by $(X_i, X_j) \in A \iff X_j \in N_{PE}(X_i, r)$. Consider the reflexivity and underlying graphs of $\mathbf{D}_n(r)$. Recall that $ij \in E_{\text{ref}}$ iff $X_j \in N_{PE}(X_i, r) \cap \Gamma_1(X_i, r)$ and $X_iX_j \in E_{\text{und}}$ iff $X_j \in N_{PE}(X_i, r) \cup \Gamma_1(X_i, r)$. See Fig. 5 for the arcs of the PE-PCD and the edges for the associated graphs with $r = 3/2$ for 7 $\mathcal{X}$ points iid uniformly generated in the triangle based on 3 $\mathcal{Y}$ points. In the general vertex random graph notation of Section 2, we have $\mathscr{V} = T(\mathcal{Y}_3)$ and $\mu = \mathcal{U}(T(\mathcal{Y}_3))$.

The edge density $\rho_n^{\text{ref}}(r) := \rho_{\text{ref}}(\mathbf{D}_n(r))$ depends on n explicitly, and on F and $N_{PE}(\cdot, r)$ implicitly. Let $p_{\text{ref}}(F, r) := \mathbf{E}[\psi_{\text{ref}}(X_1, X_2, r)]$ and $\nu_{\text{ref}}(F, r) := \mathbf{Cov}[\psi_{\text{ref}}(X_1, X_2, r), \psi_{\text{ref}}(X_1, X_3, r)]$ where $\psi_{\text{ref}}(X_1, X_2, r) = \mathbf{I}(X_2 \in N_{PE}(X_1, r) \cap \Gamma_1(X_1, r))$. Then $\mathbf{E}\left[\rho_n^{\text{ref}}(r)\right] = p_{\text{ref}}(F, r)$. By Theorem 3.5, it follows that

$$\sqrt{n}\left(\rho_n^{\text{ref}}(r) - p_{\text{ref}}(F, r)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_{\text{ref}}(F, r)) \tag{15}$$

provided $\nu_{\text{ref}}(F, r) > 0$. The asymptotic variance of $\rho_n^{\text{ref}}(r)$ is $4\,\nu_{\text{ref}}(F, r)$ and depends on only F and $N_{PE}(\cdot, r)$. Thus we need only determine $p_{\text{ref}}(F, r)$ and $\nu_{\text{ref}}(F, r)$ in order to obtain the normal

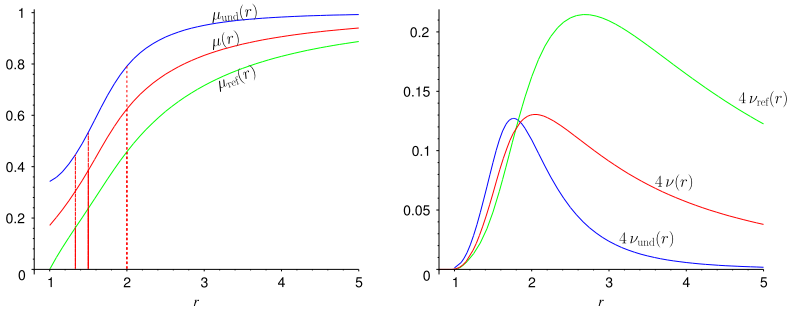

Fig. 6. Result of Theorem 4.2: asymptotic null means $p(r)$, $p_{\text{ref}}(r)$, and $p_{\text{und}}(r)$ (left) and variances $4\,\nu(r)$, $4\,\nu_{\text{ref}}(r)$, and $4\,\nu_{\text{und}}(r)$ (right) for $r \in [1, 5]$.

approximation

$$\rho_n^{\text{ref}}(r) \overset{\text{approx}}{\sim} \mathcal{N}\left(p_{\text{ref}}(F, r), \frac{4\,\nu_{\text{ref}}(F, r)}{n}\right). \tag{16}$$

The above statements hold for $\rho_n^{\text{und}}(r) := \rho_{\text{und}}(\mathbf{D}_n(r))$ also with 'ref's being replaced with 'und's and $\cap$'s with $\cup$'s.

Let F_c be a continuous distribution on $T(\mathcal{Y}_3)$ and F be any distribution on $T(\mathcal{Y}_3)$. For $r = 1$, we have $N_{PE}(x, 1) \cap \Gamma_1(x, 1) = \ell(v(x), x) \cap T(\mathcal{Y}_3)$ which has zero $\mathbb{R}^2$-Lebesgue measure. Then for F_c we have

$$\begin{aligned}\mathbf{E}\left[\rho_n^{\text{ref}}(r = 1)\right] &= \mathbf{E}\left[\psi_{\text{ref}}(X_1, X_2, r = 1)\right] = p_{\text{ref}}(F_c, r = 1)\\ &= P_{F_c}(X_2 \in N_{PE}(X_1, 1) \cap \Gamma_1(X_1, 1)) = 0.\end{aligned}$$

Similarly, $P_{F_c}(\{X_2, X_3\} \subset N_{PE}(X_1, 1) \cap \Gamma_1(X_1, 1)) = 0$. Thus, $\nu_{\text{ref}}(F_c, r = 1) = 0$. Furthermore, for $r = \infty$, $N_{PE}(x, r = \infty) \cap \Gamma_1(x, r = \infty) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$. Then

$$\begin{aligned}\mathbf{E}\left[\rho_n^{\text{ref}}(r = \infty)\right] &= \mathbf{E}\left[\psi_{\text{ref}}(X_1, X_2, \infty)\right] = p_{\text{ref}}(F, \infty)\\ &= P_F(X_2 \in N_{PE}(X_1, \infty) \cap \Gamma_1(X_1, \infty)) = P_F(X_2 \in T(\mathcal{Y}_3)) = 1.\end{aligned}$$

Similarly, $P_F(\{X_2, X_3\} \subset N_{PE}(X_1, \infty) \cap \Gamma_1(X_1, \infty)) = 1$. Hence $\nu_{\text{ref}}(F, \infty) = 0$. Therefore, the CLT result in Eq. (16) does not hold for $r = 1$ if F is continuous, and for $r = \infty$ for any F. Furthermore, $\rho_n^{\text{ref}}(r = 1) = 0$ a.s. for a continuous distribution F_c and $\rho_n^{\text{ref}}(r = \infty) = 1$ a.s. for any distribution F. For $r \in (1, \infty)$, $\psi_{\text{ref}}(X_1, X_2, r)$ tends to be high if the intersection region is large, since $\psi_{\text{ref}}(X_1, X_2, r)$ is the number of edges between vertices 1 and 2 in the reflexivity graph. In such a case, $\psi_{\text{ref}}(X_1, X_3, r)$ tends to be high also. That is, $\psi_{\text{ref}}(X_1, X_2, r)$ and $\psi_{\text{ref}}(X_1, X_3, r)$ tend to be high and low together. So, for $r \in (1, \infty)$, we have $\nu_{\text{ref}}(F, r) > 0$. See also Fig. 6 (right) and Appendix 2.

For $r = 1$, $N_{PE}(x, 1) \cup \Gamma_1(x, 1)$ has positive $\mathbb{R}^2$-Lebesgue measure. Then $P_{F_c}(\{X_2, X_3\} \subset N_{PE}(X_1, 1) \cup \Gamma_1(X_1, 1)) > 0$. Thus, $\nu_{\text{und}}(F_c, r = 1) \neq 0$. On the other hand, for $r = \infty$, $N_{PE}(X_1, \infty) \cup \Gamma_1(X_1, \infty) = T(\mathcal{Y}_3)$ for all $X_1 \in T(\mathcal{Y}_3)$. Then

$$\begin{aligned}\mathbf{E}\left[\rho_n^{\text{und}}(r = \infty)\right] &= \mathbf{E}\left[\psi_{\text{und}}(X_1, X_2, \infty)\right] = P_F(X_2 \in N_{PE}(X_1, \infty) \cup \Gamma_1(X_1, \infty))\\ &= p_{\text{und}}(F_c, \infty) = P_F(X_2 \in T(\mathcal{Y}_3)) = 1.\end{aligned}$$

Similarly, $P_F(\{X_2, X_3\} \subset N_{PE}(X_1, \infty) \cup \Gamma_1(X_1, \infty)) = 1$. Hence $\nu_{\text{und}}(F, \infty) = 0$. Therefore, the CLT result for the underlying graph does not hold for $r = \infty$ for any distribution F. Moreover $\rho_n^{\text{und}}(r = \infty) = 1$ a.s. for such F. For $r \in [1, \infty)$, since $\psi_{\text{und}}(X_1, X_2, r)$ is the number of edges in the underlying graph and tends to be high if the union region is large. In such a case, $\psi_{\text{und}}(X_1, X_3, r)$ tends to be high also. That is, $\psi_{\text{und}}(X_1, X_2, r)$ and $\psi_{\text{und}}(X_1, X_3, r)$ tend to be high and low together. So, for $r \in [1, \infty)$, we have $\nu_{\text{und}}(F, r) > 0$ for continuous F. See also Fig. 6 (right) and Appendix 3.

Remark 3.8 (*Arc Density of PE-PCDs*). Let $2\,h(X_i, X_j, r) := \mathbf{I}((X_i, X_j) \in A) + \mathbf{I}((X_j, X_i) \in A) = \mathbf{I}(X_j \in N_{PE}(X_i, r)) + \mathbf{I}(X_i \in N_{PE}(X_j, r))$ for $i \neq j$ and the arc density $\rho_n(r) := \rho(\mathbf{D}_n(r))$. Let $p_a(r) := \mathbf{E}[\rho_n(r)]$

and $4\,\nu(r) := \mathbf{Cov}\,[h(X_1, X_2, r), h(X_1, X_3, r)]$. By Theorem 3.2, we have

$$\sqrt{n}\,(\rho_n(r) - p_a(r)) \xrightarrow{\mathcal{L}} \mathcal{N}\,(0, 4\,\nu(r)) \tag{17}$$

provided $\nu(r) > 0$. The explicit forms of asymptotic mean $p(r)$ and variance $4\,\nu(r)$ for uniform data are provided in [10]. □

4. Asymptotic distribution of edge density for uniform data

Let $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ for $i = 1, 2, \ldots, n$, where $\mathcal{U}(T(\mathcal{Y}_3))$ is the uniform distribution on the triangle $T(\mathcal{Y}_3)$. We first present a "geometry invariance" result which will simplify our subsequent analysis by allowing us to consider the special case of the equilateral triangle. Let $\rho_n^{\text{ref}}(r) := \rho_{\text{ref}}(\mathcal{U}(T(\mathcal{Y}_3)), r)$ and $\rho_n^{\text{und}}(r) := \rho_{\text{und}}(\mathcal{U}(T(\mathcal{Y}_3)), r)$.

Theorem 4.1 (*Geometry Invariance*). *Let $\mathcal{Y}_3 = \{\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3\} \subset \mathbb{R}^2$ be three non-collinear points. For $i = 1, 2, \ldots, n$, let $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$. Then for any $r \in [1, \infty]$, the distribution of $\rho_n^{ref}(r)$ and $\rho_n^{und}(r)$ is independent of $\mathcal{Y}_3$, and hence the geometry of $T(\mathcal{Y}_3)$.*

Proof. For $r = 1$, $\rho_n^{\text{ref}}(1)$ is degenerate, i.e., $\rho_n^{\text{ref}}(1) = 0$ a.s. for any continuous F in $T(\mathcal{Y}_3)$ and $\rho_n^{\text{ref}}(\infty) = 1$ a.s. for any distribution F in $T(\mathcal{Y}_3)$ regardless of the geometry of $T(\mathcal{Y}_3)$. Similarly, $\rho_n^{\text{und}}(\infty) = 1$ a.s. for any distribution F in $T(\mathcal{Y}_3)$ regardless of the geometry of $T(\mathcal{Y}_3)$. Hence geometry invariance follows for these r values for uniform data in $T(\mathcal{Y}_3)$ as well. A composition of translation, rotation, reflections, and scaling will take any given triangle $T_o = T(\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3)$ to the "basic" triangle $T_b = T((0, 0), (1, 0), (c_1, c_2))$ with $0 < c_1 \leq 1/2$, $c_2 > 0$ and $(1 - c_1)^2 + c_2^2 \leq 1$, preserving uniformity. The transformation $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\Phi(u, v) = \left(u + \frac{1-2c_1}{2c_2}\, v,\ \frac{\sqrt{3}}{2c_2}\, v\right)$ takes T_b to the equilateral triangle $T_e = T\left((0, 0), (1, 0), \left(1/2, \sqrt{3}/2\right)\right)$. Investigation of the Jacobian shows that Φ also preserves uniformity. Furthermore, the composition of Φ with the rigid motion transformations and scaling maps the boundary of the original triangle T_o to the boundary of the equilateral triangle T_e, the median lines of T_o to the median lines of T_e, and lines parallel to the edges of T_o to lines parallel to the edges of T_e. (A median line in a triangle is the line joining a vertex with the center of mass.) Since the joint distribution of any collection of the $\psi_{\text{ref}}(X_i, X_j, r)$ and $\psi_{\text{und}}(X_i, X_j, r)$ involves only probability content of unions and intersections of regions bounded by precisely such lines, and the probability content of such regions is preserved since uniformity is preserved, the desired result follows for $r \in (1, \infty)$ for the reflexivity graphs and for $r \in [1, \infty)$ for the underlying graphs. ■

Based on Theorem 4.1, for our proportional-edge proximity map and the uniform data, we may assume that $T(\mathcal{Y}_3)$ is a standard equilateral triangle, T_e, with vertices $\mathcal{Y}_3 = \left\{(0, 0), (1, 0), \left(1/2, \sqrt{3}/2\right)\right\}$, henceforth.

In the case of the (proportional-edge proximity map, uniform data) pair, the asymptotic distribution of $\rho_n^{\text{ref}}(r)$ and $\rho_n^{\text{und}}(r)$ can be derived as a function of r by detailed geometric computations. Recall that $p_{\text{ref}}(r) = \mathbf{E}\,[\psi_{\text{ref}}(X_1, X_2, r)] = P(X_2 \in N_{PE}(X_1, r) \cap \Gamma_1(X_1, r))$ and $p_{\text{und}}(r) = \mathbf{E}\,[\psi_{\text{und}}(X_1, X_2, r)] = P(X_2 \in N_{PE}(X_1, r) \cup \Gamma_1(X_1, r))$ are the edge probabilities in the reflexivity and underlying graphs, respectively.

Theorem 4.2. *For $r \in (1, \infty)$,*

$$\sqrt{n}\left(\rho_n^{ref}(r) - p_{ref}(r)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_{ref}(r))$$

and for $r \in [1, \infty)$,

$$\sqrt{n}\left(\rho_n^{und}(r) - p_{und}(r)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\,\nu_{und}(r))$$

where the asymptotic means are

$$p_{ref}(r) = \begin{cases} \dfrac{(1-r)(5\,r^5 - 148\,r^4 + 245\,r^3 - 178\,r^2 - 232\,r + 128)}{54\,r^2(r+2)(r+1)} \\ \quad \textit{for } r \in [1, 4/3), \\ \dfrac{-101\,r^5 + 801\,r^4 - 1302\,r^3 + 732\,r^2 + 536\,r - 672}{216\,r(r+2)(r+1)} \\ \quad \textit{for } r \in [4/3, 3/2), \\ \dfrac{r^8 - 13\,r^7 + 30\,r^6 + 148\,r^5 - 448\,r^4 + 264\,r^3 + 288\,r^2 - 368\,r + 96}{8\,r^4(r+2)(r+1)} \\ \quad \textit{for } r \in [3/2, 2), \\ \dfrac{(r^3 + 3\,r^2 - 2 + 2\,r)(1-r)^2}{r^4(r+1)} \quad \textit{for } r \in [2, \infty) \end{cases} \tag{18}$$

and

$$p_{und}(r) = \begin{cases} \dfrac{47\,r^6 - 195\,r^5 + 860\,r^4 - 846\,r^3 - 108\,r^2 + 720\,r - 256}{108\,r^2(r+2)(r+1)} \\ \quad \textit{for } r \in [1, 4/3), \\ \dfrac{175\,r^5 - 579\,r^4 + 1450\,r^3 - 732\,r^2 - 536\,r + 672}{216\,r\,(r+2)(r+1)} \\ \quad \textit{for } r \in [4/3, 3/2), \\ \dfrac{-3\,r^8 + 7\,r^7 + 30\,r^6 - 84\,r^5 + 264\,r^4 - 304\,r^3 - 144\,r^2 + 368\,r - 96}{8\,r^4(r+2)(r+1)} \\ \quad \textit{for } r \in [3/2, 2), \\ \dfrac{r^5 + r^4 - 6\,r + 2}{r^4(r+1)} \quad \textit{for } r \in [2, \infty), \end{cases} \tag{19}$$

and the asymptotic variances are

$$\nu_{ref}(r) = \sum_{i=1}^{11} \vartheta_i^{ref}(r)\,\mathbf{I}(\mathcal{I}_i) \textit{ and } \nu_{und}(r) = \sum_{i=1}^{11} \vartheta_i^{und}(r)\,\mathbf{I}(\mathcal{I}_i). \tag{20}$$

The explicit forms of $\vartheta_i^{\text{ref}}(r)$ and $\vartheta_i^{\text{und}}(r)$ are provided in Appendix Sections 2 and 3, and the derivations of $p_{\text{ref}}(r)$, $\nu_{\text{ref}}(r)$, $p_{\text{und}}(r)$, and $\nu_{\text{und}}(r)$ are provided in [7].

The expectation $\mathbf{E}[\psi_{\text{ref}}(X_1, X_2, r)] = p_{\text{ref}}(r)$ is as in Eq. (18); and $\mathbf{E}[\psi_{\text{und}}(X_1, X_2, r)] = p_{\text{und}}(r)$ is as in Eq. (19). See Fig. 6 for the plots of the asymptotic means and variances. Some values of note on this plot are $p(1) = 37/216$, $p_{\text{ref}}(1) = 0$, and $p_{\text{und}}(1) = 37/108$, $\lim_{r\to\infty} p(r) = \lim_{r\to\infty} p_{\text{ref}}(r) = \lim_{r\to\infty} p_{\text{und}}(r) = 1$, $4\,\nu_{\text{ref}}(r = 1) = 0$ and $\lim_{r\to\infty} 4\,\nu_{\text{ref}}(r) = 0$, $4\,\nu_{\text{und}}(r = 1) = 1/3240$ and $\lim_{r\to\infty} 4\,\nu_{\text{und}}(r) = 0$, and $\text{argsup}_{r\in[1,\infty]} 4\,\nu(r) \approx 2.045$ with $\sup_{r\in[1,\infty]} 4\,\nu(r) \approx .1305$, $\text{argsup}_{r\in[1,\infty]} 4\,\nu_{\text{ref}}(r) \approx 2.69$ with $\sup_{r\in[1,\infty]} 4\,\nu_{\text{ref}}(r) \approx .0537$, $\text{argsup}_{r\in[1,\infty]} 4\,\nu_{\text{und}}(r) \approx 1.765$ with $\sup_{r\in[1,\infty]} 4\,\nu_{\text{und}}(r) \approx .0318$. Notice that $p_{\text{ref}}(r = 1) = 0$ and $\lim_{r\to\infty} p_{\text{ref}}(r) = 1$ (at rate $O\left(r^{-1}\right)$); and $p_{\text{und}}(r = 1) = 37/108$ and $\lim_{r\to\infty} p_{\text{und}}(r) = 1$ (at rate $O\left(r^{-1}\right)$).

By construction of the reflexivity and underlying graphs, there is a natural ordering of the means of arc and edge densities.

Lemma 4.3. *The means of the edge densities and arc density (i.e., the edge and arc probabilities) have the following ordering: $p_{ref}(r) < p(r) < p_{und}(r)$ for all $r \in [1, \infty)$. Furthermore, for $r = \infty$, we have $p_{ref}(r) = p(r) = p_{und}(r) = 1$.*

Proof. Recall that $p_{\text{ref}}(r) = \mathbf{E}[\rho_n^{\text{ref}}(r)] = P(X_2 \in N_{PE}(X_1, r) \cap \Gamma_1(X_1, r))$, $p(r) = \mathbf{E}[\rho_n(r)] = P(X_2 \in N_{PE}(X_1, r))$, and $p_{\text{und}}(r) = \mathbf{E}[\rho_n^{\text{und}}(r)] = P(X_2 \in N_{PE}(X_1, r) \cup \Gamma_1(X_1, r))$. Additionally, $[N_{PE}(X_1, r) \cap \Gamma_1(X_1, r)] \subseteq N_{PE}(X_1, r) \subseteq [N_{PE}(X_1, r) \cup \Gamma_1(X_1, r)]$ with probability 1 for all $r \geq 1$ with equality holding for $r = \infty$ only. Then the desired results follow (see also Fig. 6). ■

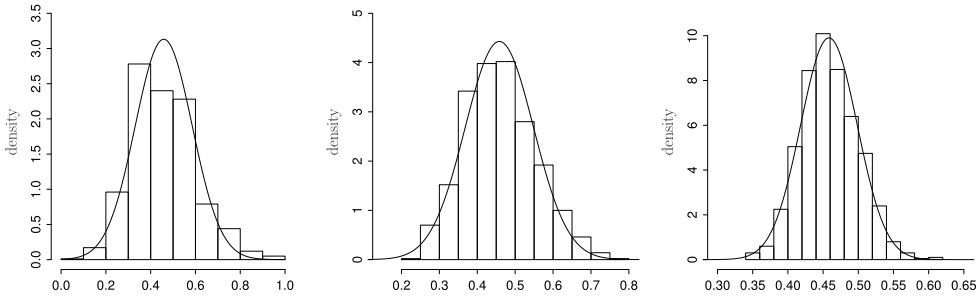

Fig. 7. Depicted are $\rho_n^{\text{ref}}(2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{11}{24}, \frac{58901}{362880\,n}\right)$ for $n = 10,\ 20,\ 100$ (left to right). Histograms are based on 1000 Monte Carlo replicates. Solid lines are the corresponding normal densities. Notice that the axes are differently scaled.

Note that the above lemma holds for all X_i that has a continuous distribution on $T(\mathcal{Y}_3)$. There is also a stochastic ordering for the edge and arc densities as follows.

Theorem 4.4. *For sufficiently small r, $\rho_n^{ref}(r) <^{st} \rho_n(r) <^{st} \rho_n^{und}(r)$ as $n \to \infty$ where $<^{st}$ stands for "stochastically smaller than".*

Proof. Above we have proved that $p_{\text{ref}}(r) < p(r) < p_{\text{und}}(r)$ for all $r \in [1, \infty)$. For small r (i.e., $r \lesssim 1.8$) the asymptotic variances have the same ordering, i.e., $4\,\nu_{\text{ref}}(r) < 4\,\nu(r) < 4\,\nu_{\text{und}}(r)$. Since $\rho_n^{\text{ref}}(r)$, $\rho_n(r)$, and $\rho_n^{\text{und}}(r)$ are asymptotically normal, then the desired result follows (see also Fig. 6). ■

To illustrate the limiting distribution, for example, $r = 2$ yields

$$\frac{\sqrt{n}(\rho_n^{\text{ref}}(2) - p_{\text{ref}}(2))}{\sqrt{4\,\nu_{\text{ref}}(2)}} = \sqrt{\frac{362880\,n}{58901}}\left(\rho_n^{\text{ref}}(2) - \frac{11}{24}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

and

$$\frac{\sqrt{n}(\rho_n^{\text{und}}(2) - p_{\text{und}}(2))}{\sqrt{4\,\nu_{\text{und}}(2)}} = \sqrt{\frac{120960\,n}{13189}}\left(\rho_n^{\text{und}}(2) - \frac{19}{24}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1);$$

or equivalently,

$$\rho_n^{\text{ref}}(2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{11}{24}, \frac{58901}{362880\,n}\right) \quad \text{and} \quad \rho_n^{\text{und}}(2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{19}{24}, \frac{13189}{120960\,n}\right).$$

We assess the accuracy of the normal approximation for finite sample data based on Monte Carlo simulations. We generate $n\mathcal{X}$ points independently uniformly in the standard equilateral triangle T_e. For each data set generated, we calculate the edge density of the reflexivity and underlying graphs based on the PE-PCD with $r = 2$. We replicate the above process $N_{mc} = 1000$ times for each of $n = 10, 20$, and 100. We plot the histograms of the edge density values of these graphs using the simulated data and the corresponding (asymptotic) normal curves in Figs. 7 and 8, respectively. Notice that, for $r = 2$, the normal approximation is accurate even for small n although skewness may be indicated for $n = 10$ in both reflexivity and underlying graph cases. In particular, the edge density is skewed right (resp. left) for the reflexivity (resp. underlying) graph.

Recall that the upper bounds of the sharpest rates of convergence provided in Remark 3.6 is $K\frac{p_a}{\sqrt{n\,\nu_a^3}}$ for a vertex random digraph. To compare the rates of convergence bounds, we consider the version of these upper bounds scaled by $\sqrt{n}/K$, i.e., $\frac{p_a(r)}{\sqrt{\nu_a^3(r)}}$ for the PE-PCD, $\frac{p_{\text{ref}}(r)}{\sqrt{\nu_{\text{ref}}^3(r)}}$, and $\frac{p_{\text{und}}(r)}{\sqrt{\nu_{\text{und}}^3(r)}}$ for the reflexivity and underlying graphs, respectively. Then, the smaller this bound, the faster the rate of convergence to normality. We plot these rates in Fig. 9 for $r \in [1, 5]$ in two panels for better visualization. Notice that sharpest rate occurs at $r \approx 1.88$ for the arc density of PE-PCDs, and at $r \approx 2.35$ (resp. $r \approx 1.69$) for the edge density of reflexivity (resp. underlying) graph. Furthermore, the rate of convergence is

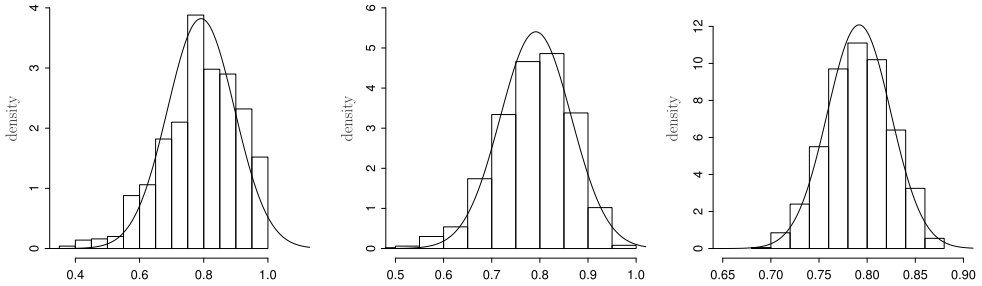

Fig. 8. Depicted are $\rho_n^{\text{und}}(2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{19}{24}, \frac{13189}{120960\,n}\right)$ for $n = 10, 20, 100$ (left to right). Histograms are based on 1000 Monte Carlo replicates. Solid lines are the corresponding normal densities. Notice that the axes are differently scaled.

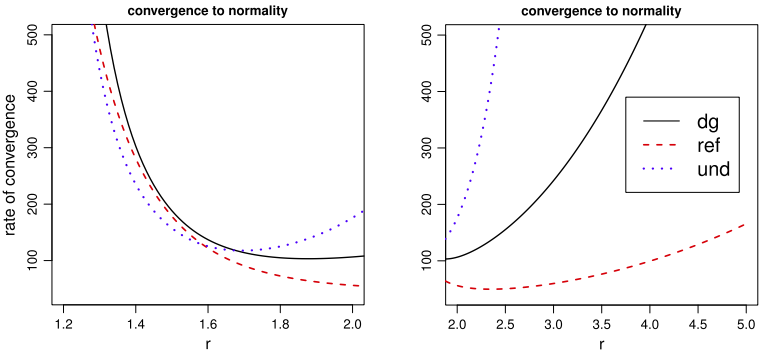

Fig. 9. The upper bounds for the convergence rates (multiplied with $\sqrt{n}/K$) for the arc density of PE-PCD and the edge densities of the associated graphs. The plots are divided into two panels for better visualization: left one is for the range of $r \in [1.2, 2]$ and right one is for $r \in [2, 5]$. Notice that x-axes are differently scaled. "dg" stands for the arc density of the PE-PCDs, "ref" (resp. "und") for the reflexivity (resp. underlying) graph.

lower for the edge density of reflexivity and underlying graphs at each r value compared to the arc density of PE-PCDs. In particular, for $r \lessapprox 1.59$, the rate is sharpest for the underlying graph, while for $r \gtrapprox 1.59$, the rate is sharpest for the reflexivity graph. For the entire range of the r values, the sharpest rate occurs for the reflexivity graph at $r \approx 2.35$.

The finite sample variance and skewness may be derived analytically in much the same way as was $4\,\nu_{\text{ref}}(r)$ (resp. $4\,\nu_{\text{und}}(r)$) for the asymptotic variance. In fact, the exact distribution of $\rho_n^{\text{ref}}(r)$ (resp. $\rho_n^{\text{und}}(r)$) is, in principle, available by successively conditioning on the values of the X_i. Alas, while the joint distribution of $\psi_{\text{ref}}(X_1, X_2, r)$, $\psi_{\text{ref}}(X_1, X_3, r)$ (resp. $\psi_{\text{und}}(X_1, X_2, r)$, $\psi_{\text{und}}(X_1, X_3, r)$) is available (see Figs. 10 and 11), the joint distribution of $\{\psi_{\text{ref}}(X_i, X_j, r)\}_{1\le i<j\le n}$ (resp. $\{\psi_{\text{und}}(X_i, X_j, r)\}_{1\le i<j\le n}$), and hence the calculation for the exact distribution of $\rho_n^{\text{ref}}(r)$ (resp. $\rho_n^{\text{und}}(r)$), is extraordinarily tedious and lengthy even for small values of n. Hence, this aspect is not pursued any further.

Let $\gamma_n(r)$ be the domination number of the PE-PCD based on $\mathcal{X}_n$ which is a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. Additionally, let $\gamma_n^{\text{ref}}(r)$ and $\gamma_n^{\text{und}}(r)$ be the domination numbers of the reflexivity and underlying graphs based on the PE-PCD, respectively. Then we have the following stochastic ordering for the domination numbers of the PE-PCD and the associated graphs.

Theorem 4.5. *For all $r \in [1, \infty)$ and $n > 1$, $\gamma_n^{und}(r) <^{st} \gamma_n(r) <^{st} \gamma_n^{ref}(r)$.*

Proof. For all $x \in T(\mathcal{Y}_3)$, we have $[N_{PE}(x, r) \cap \Gamma_1(x, r)] \subseteq N_{PE}(x, r) \subseteq [N_{PE}(x, r) \cup \Gamma_1(x, r)]$. For $X \sim \mathcal{U}(T(\mathcal{Y}_3))$, we have $[N_{PE}(X, r) \cap \Gamma_1(X, r)] \subsetneq N_{PE}(X, r) \subsetneq [N_{PE}(X, r) \cup \Gamma_1(X, r)]$ a.s. Moreover, $\gamma_n(r) = 1$ iff $\mathcal{X}_n \subset N_{PE}(X_i, r)$ for some i; $\gamma_n^{\text{ref}}(r) = 1$ iff $\mathcal{X}_n \subset N_{PE}(X_i, r) \cap \Gamma_1(X_i, r)$ for some i; and $\gamma_n^{\text{und}}(r) = 1$ iff $\mathcal{X}_n \subset N_{PE}(X_i, r) \cup \Gamma_1(X_i, r)$ for some i. So it follows that $P(\gamma_n^{\text{ref}}(r) = 1) < P(\gamma_n(r) = 1) < P(\gamma_n^{\text{und}}(r) = 1)$. Similarly, for all $x, y \in T(\mathcal{Y}_3)$, we have $\Big([N_{PE}(x, r) \cap \Gamma_1(x, r)] \cup [N_{PE}(y, r)$

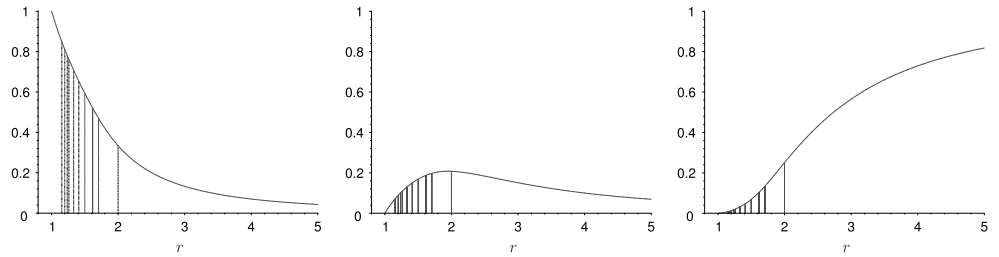

Fig. 10. The plots for the joint distribution of $\psi_{\text{ref}}(X_1, X_2, r)$, $\psi_{\text{ref}}(X_1, X_3, r)$ for $r \in [1, 5]$. Plotted are $P(\psi_{\text{ref}}(X_1, X_2, r), \psi_{\text{ref}}(X_1, X_3, r) = (0, 0))$ (left), $P(\psi_{\text{ref}}(X_1, X_2, r), \psi_{\text{ref}}(X_1, X_3, r) = (1, 0)) = P(\psi_{\text{ref}}(X_1, X_2, r), \psi_{\text{ref}}(X_1, X_3, r) = (0, 1))$ (middle), and $P(\psi_{\text{ref}}(X_1, X_2, r), \psi_{\text{ref}}(X_1, X_3, r) = (1, 1))$ (right).

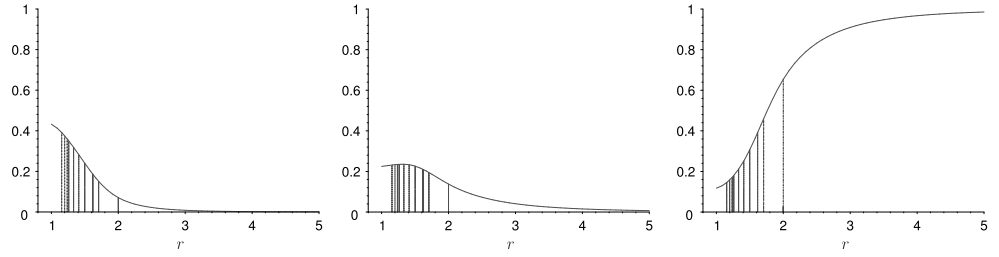

Fig. 11. The plots for the joint distribution of $\psi_{\text{und}}(X_1, X_2, r)$, $\psi_{\text{und}}(X_1, X_3, r)$ for $r \in [1, 5]$. Plotted are $P(\psi_{\text{und}}(X_1, X_2, r), \psi_{\text{und}}(X_1, X_3, r) = (0, 0))$ (left), $P(\psi_{\text{und}}(X_1, X_2, r), \psi_{\text{und}}(X_1, X_3, r) = (1, 0)) = P(\psi_{\text{und}}(X_1, X_2, r), \psi_{\text{und}}(X_1, X_3, r) = (0, 1))$ (middle), and $P(\psi_{\text{und}}(X_1, X_2, r), \psi_{\text{und}}(X_1, X_3, r) = (1, 1))$ (right).

$\cap\, \Gamma_1(y, r)]\Big) \subseteq (N_{PE}(x, r) \cup N_{PE}(y, r)) \subseteq \Big([N_{PE}(x, r) \cup \Gamma_1(x, r)] \cup [N_{PE}(y, r) \cup \Gamma_1(y, r)]\Big)$. For $X, Y \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$, we have $\Big([N_{PE}(X, r) \cap \Gamma_1(X, r)] \cup [N_{PE}(Y, r) \cap \Gamma_1(Y, r)]\Big) \subsetneq (N_{PE}(X, r) \cup N_{PE}(Y, r)) \subsetneq \Big([N_{PE}(X, r) \cup \Gamma_1(X, r)] \cup [N_{PE}(Y, r) \cup \Gamma_1(Y, r)]\Big)$ a.s. Moreover, $\gamma_n(r) \leq 2$ iff $\mathcal{X}_n \subset N_{PE}(X_i, r) \cup N_{PE}(X_j, r)$ for some $i \neq j$; $\gamma_n^{\text{ref}}(r) \leq 2$ iff $\mathcal{X}_n \subset \Big([N_{PE}(X_i, r) \cap \Gamma_1(X_i, r)] \cup [N_{PE}(X_j, r) \cap \Gamma_1(X_j, r)]\Big)$ for some $i \neq j$; and $\gamma_n^{\text{und}}(r) \leq 2$ iff $\mathcal{X}_n \subset \Big([N_{PE}(X_i, r) \cup \Gamma_1(X_i, r)] \cup [N_{PE}(X_j, r) \cup \Gamma_1(X_j, r)]\Big)$ for some $i \neq j$. So it follows that $P(\gamma_n^{\text{ref}}(r) \leq 2) < P(\gamma_n(r) \leq 2) < P(\gamma_n^{\text{und}}(r) \leq 2)$. Since $P(\gamma_n(r) \leq 3) = 1$ [8], it follows that $P(\gamma_n^{\text{und}}(r) \leq 3) = 1$ also holds since $P(\gamma_n(r) \leq 3) \leq P(\gamma_n^{\text{und}}(r) \leq 3)$. Hence the desired stochastic ordering follows. ■

Note the stochastic ordering in the above theorem holds for any continuous distribution F with support being in $T(\mathcal{Y}_3)$. For $r = \infty$, we have $\gamma_n^{\text{und}}(r) = \gamma_n(r) = \gamma_n^{\text{ref}}(r) = 1$ a.s.

5. Multiple triangle case

Suppose $\mathcal{Y}_m$ is a finite set of $m > 3$ points in $\mathbb{R}^2$. Consider the Delaunay triangulation (assumed to exist) of $\mathcal{Y}_m$. Let T_i denote the ith Delaunay triangle, J_m denote the number of triangles, and $C_H(\mathcal{Y}_m)$ denote the convex hull of $\mathcal{Y}_m$. For $X_i \overset{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}_m))$, $i = 1, 2, \ldots, n$, we construct the PE-PCD, $\mathbf{D}_{n,m}(r)$, using $N_{PE}(\cdot, r)$ as described in Section 3.4, where for $X_i \in T_j$, the three points in $\mathcal{Y}_m$ defining the Delaunay triangle T_j are denoted as $\mathcal{Y}_{[j]}$. Here we have $\mathscr{V} = C_H(\mathcal{Y}_m)$ and $\mu = \mathcal{U}(C_H(\mathcal{Y}_m))$ in the vertex random graph notation of Section 2. See Fig. 12 for the arcs of the PE-PCD and the edges for the reflexivity and underlying graphs with $r = 3/2$ and the $\mathcal{X}$ and $\mathcal{Y}$ points from Figs. 1 and 2. We investigate the edge densities of the graphs based on the PE-PCD. We consider various versions of the edge density in the multiple triangle case. Observe that by construction, the PE-PCD has J_m

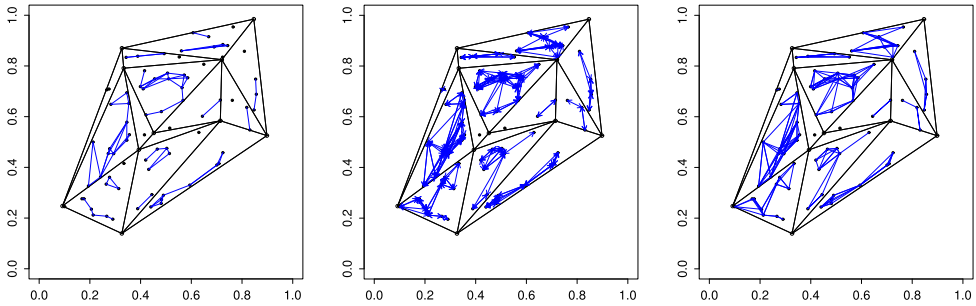

Fig. 12. Presented in the middle are the arcs for the PE-PCD with $r = 3/2$ for the 77 $\mathcal{X}$ points from Fig. 2 and the 10 $\mathcal{Y}$ points from Fig. 1. Plotted in the left are the edges for the associated reflexivity graph, and right are for the underlying graph.

components, each of which might be void, connected, or disconnected, since the arcs do not cross the boundaries of the Delaunay triangles, and neither do the edges of the reflexivity and underlying graphs. Furthermore, by geometry invariance for uniform data (see Theorem 4.1), the arc probability for the PE-PCD, and edge probability for the underlying graphs (and the reflexivity graphs) are same for each triangle and equal to the ones in the standard equilateral triangle. The same holds for the asymptotic variance of arc and edge density values. Hence, for the multiple triangle case, we use these parameters of the asymptotic normality from the one-triangle case.

5.1. First version of edge density in the multiple triangle case

For $J_m > 1$, as in Section 3.5, let $\rho^{\text{ref}}_{l,n}(r) = 2\,|E_{\text{ref}}|\,/(n\,(n-1))$ and $\rho^{\text{und}}_{l,n}(r) = 2\,|E_{\text{und}}|\,/(n\,(n-1))$. Let $E^{\text{ref}}_{[i]}$ be the set of edges and $\rho^{\text{ref}}_{[i]}(r)$ be the edge density for triangle i in the reflexivity graph, and $E^{\text{und}}_{[i]}$ and $\rho^{\text{und}}_{[i]}(r)$ be similarly defined for the underlying graph. Then clearly we have $|E_{\text{ref}}| = \sum_{i=1}^{J_m} \left|E^{\text{ref}}_{[i]}\right|$ and $|E_{\text{und}}| = \sum_{i=1}^{J_m} \left|E^{\text{und}}_{[i]}\right|$. Let n_i be the number of X points in T_i for $i = 1, 2, \ldots, J_m$. Letting $w_i = A(T_i)/A(C_H(\mathcal{Y}_m))$ with $A(\cdot)$ being the area function, we obtain the following as a corollary to Theorem 4.2.

Corollary 5.1. *For $r \in (1, \infty)$, the asymptotic distribution for $\rho^{ref}_{l,n}(r)$ conditional on $\mathcal{Y}_m$ is given by*

$$\sqrt{n}\left(\rho^{\text{ref}}_{l,n}(r) - \widetilde{p}_{ref}(m,r)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\,\widetilde{v}_{ref}(m,r)\right), \tag{21}$$

as $n \to \infty$, where $\widetilde{p}_{ref}(m,r) = p_{ref}(r)\left(\sum_{i=1}^{J_m} w_i^2\right)$ and

$$\widetilde{v}_{ref}(m,r) = v_{ref}(r)\left(\sum_{i=1}^{J_m} w_i^3\right) + p^2_{ref}(r)\left(\sum_{i=1}^{J_m} w_i^3 - \left(\sum_{j=1}^{J_m} w_i^2\right)^2\right)$$

with $p_{ref}(r)$ and $v_{ref}(r)$ being as in Eqs. (18) and (20), respectively. The asymptotic distribution of $\rho^{und}_{l,n}(r)$ with $r \in [1, \infty)$ is similar.

The proof is provided in Appendix 4. By an appropriate application of the Jensen's inequality, we see that $\sum_{i=1}^{J_m} w_i^3 \geq \left(\sum_{i=1}^{J_m} w_i^2\right)^2$. So the covariance above is zero iff $v_{\text{ref}}(r) = 0$ and $\sum_{i=1}^{J_m} w_i^3 = \left(\sum_{i=1}^{J_m} w_i^2\right)^2$, so asymptotic normality may hold even though $v_{\text{ref}}(r) = 0$. That is, $\rho^{\text{ref}}_{l,n}(r)$ has the asymptotic normality for $r \in \{1, \infty\}$ also provided that $\sum_{i=1}^{J_m} w_i^3 > \left(\sum_{i=1}^{J_m} w_i^2\right)^2$. The same holds for the underlying graph for $r = \infty$.

5.2. Other versions of edge density in the multiple triangle case

Let $\Xi_n^{\text{ref}}(r) := \sum_{i=1}^{J_m} \frac{n_i\,(n_i-1)}{n\,(n-1)} \rho_{[i]}^{\text{ref}}(r)$ and $\Xi_n^{\text{und}}(r) := \sum_{i=1}^{J_m} \frac{n_i\,(n_i-1)}{n\,(n-1)} \rho_{[i]}^{\text{und}}(r)$. Then $\Xi_n^{\text{ref}}(r) = \rho_{I,n}^{\text{ref}}(r)$, since $\Xi_n^{\text{ref}}(r) = \sum_{i=1}^{J_m} \frac{n_i\,(n_i-1)}{n\,(n-1)} \rho_{[i]}^{\text{ref}}(r) = \frac{\sum_{i=1}^{J_m} 2\left|E_{[i]}^{\text{ref}}\right|}{n\,(n-1)} = \frac{2\,|E_{\text{ref}}|}{n\,(n-1)} = \rho_{I,n}^{\text{ref}}(r)$. Similarly, we have $\Xi_n^{\text{und}}(r) = \rho_n^{\text{und}}(r)$.

Furthermore, let $\widehat{\Xi}_n^{\text{ref}} := \sum_{i=1}^{J_m} w_i^2\, \rho_{[i]}^{\text{ref}}(r)$ where w_i is as in Section 5.1. So $\widehat{\Xi}_n^{\text{ref}}$ a linear combination of the $\rho_{[i]}^{\text{ref}}(r)$ values. Since the $\rho_{[i]}^{\text{ref}}(r)$ are asymptotically independent, $\widehat{\Xi}_n^{\text{ref}}(r)$ and $\rho_{I,n}^{\text{ref}}(r)$ are asymptotically normal; i.e., for large n their distribution is approximately $\mathcal{N}\left(\widetilde{p}_{\text{ref}}(m,r), 4\,\widetilde{\nu}_{\text{ref}}(m,r)/n\right)$. A similar result holds for the underlying graph case.

In Section 5.1, the denominator of $\rho_{I,n}^{\text{ref}}(r)$ has $n(n-1)/2$ as the maximum number of edges possible. However, by definition, given the n_i values, we can have a graph with at most J_m complete components, each with order n_i for $i = 1, 2, \ldots, J_m$. Then the maximum number of edges possible is $n_t := \sum_{i=1}^{J_m} n_i\,(n_i-1)/2$ which suggests another version of edge density, namely, $\rho_{II,n}^{\text{ref}}(r) := \frac{|E_{\text{ref}}|}{n_t}$. Then $\rho_{II,n}^{\text{ref}}(r) = \frac{\sum_{i=1}^{J_m}\left|E_{[i]}^{\text{ref}}\right|}{n_t} = \sum_{i=1}^{J_m} \frac{n_i\,(n_i-1)}{2\,n_t}\, \rho_{[i]}^{\text{ref}}(r)$. Since $\frac{n_i\,(n_i-1)}{2\,n_t} \geq 0$ for each i, and $\sum_{i=1}^{J_m} \frac{n_i\,(n_i-1)}{2\,n_t} = 1$, it follows that $\rho_{II,n}^{\text{ref}}(r)$ is a mixture of the $\rho_{[i]}^{\text{ref}}(r)$. Then $\mathbf{E}\left[\rho_{II,n}^{\text{ref}}(r)\right] = p_{\text{ref}}(r)$. A similar result holds for the underlying graph as well.

Theorem 5.2. *The asymptotic distribution for $\rho_{II,n}^{ref}(r)$ conditional on $\mathcal{Y}_m$ for $r \in (1,\infty)$ is given by*

$$\sqrt{n}\left(\rho_{II,n}^{ref}(r) - p_{ref}(r)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\, 4\,\breve{\nu}_{ref}(m,r)\right), \tag{22}$$

as $n \to \infty$, where $\breve{\nu}_{ref}(m,r) = \nu_{ref}(r)\left(\sum_{i=1}^{J_m} w_i^3\right) / \left(\sum_{i=1}^{J_m} w_i^2\right)^2$ with $p_{ref}(r)$ and $\nu_{ref}(r)$ being as in Eqs. (18) and (20), respectively. The asymptotic distribution of $\rho_{II,n}^{und}(r)$ with $r \in [1,\infty)$ is similar.

The proof is provided in Appendix 5. Notice that the covariance $\breve{\nu}_{\text{ref}}(m,r)$ is zero iff $\nu_{\text{ref}}(r) = 0$. The underlying graph case is similar.

Remark 5.3 (*Comparison of Versions of Edge Density in the Multiple Triangle Case*)**.** Among the versions of the edge density we considered, we have $\Xi_n^{\text{ref}}(r) = \rho_{I,n}^{\text{ref}}(r)$ for all $n > 1$, and $\widehat{\Xi}_n^{\text{ref}}(r)$ and $\rho_{I,n}^{\text{ref}}(r)$ are asymptotically equivalent (i.e., they have the same asymptotic distribution). However, $\rho_{I,n}^{\text{ref}}(r)$ and $\rho_{II,n}^{\text{ref}}(r)$ do not have the same distribution for finite or infinite n. But we have $\rho_{I,n}^{\text{ref}}(r) = \frac{2\,n_t}{n(n-1)} \rho_{II,n}^{\text{ref}}(r)$ and since $\sum_{i=1}^{J_m} w_i^2 < 1$, it follows that $\widetilde{p}_{\text{ref}}(m,r) < p_{\text{ref}}(r)$. Furthermore, since $\frac{2\,n_t}{n(n-1)} = \sum_{i=1}^{J_m} \frac{n_i(n_i-1)}{n(n-1)} \longrightarrow \sum_{i=1}^{J_m} w_i^2$ as $n_i \to \infty$, we have $\lim_{n_i\to\infty} \mathbf{Var}\left[\sqrt{n}\rho_{I,n}^{\text{ref}}(r)\right] = \left(\sum_{i=1}^{J_m} w_i^2\right)^2 \left(\lim_{n_i\to\infty} \mathbf{Var}\left[\sqrt{n}\rho_{II,n}^{\text{ref}}(r)\right]\right)$. Hence $\widetilde{\nu}_{\text{ref}}(m,r) = \left(\sum_{i=1}^{J_m} w_i^2\right)^2 \breve{\nu}_{\text{ref}}(m,r)$. Since $\left(\sum_{i=1}^{J_m} w_i^2\right)^2 < 1$, it follows that $\breve{\nu}_{\text{ref}}(m,r) > \widetilde{\nu}_{\text{ref}}(m,r)$ which implies that the rate of convergence to normality is sharper for $\rho_{II,n}^{\text{ref}}(r)$. Moreover, asymptotic normality might hold for $\rho_{I,n}^{\text{ref}}(r)$ or $\widehat{\Xi}_n^{\text{ref}}(r)$ even if $\nu_{\text{ref}}(r) = 0$. The same conclusions hold for the edge density of the underlying graph as well. □

5.3. Extension to higher dimensions

The extension to $\mathbb{R}^d$ with $d > 2$ is straightforward. Let $\mathcal{Y}_{d+1} = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_{d+1}\}$ be $d+1$ non-coplanar points. 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 $(d-1)$. For $r \in [1,\infty]$, define the proportional-edge proximity map as follows. Given a point x in $\mathcal{S}(\mathcal{Y}_{d+1})$, let $\mathsf{y} := \arg\min_{\mathsf{y}\in\mathcal{Y}_{d+1}} \text{volume}(Q_{\mathsf{y}}(x))$ where $Q_{\mathsf{y}}(x)$ is the polytope with vertices being the $d\,(d+1)/2$

midpoints of the edges, the vertex y and x. That is, the vertex region for vertex v is the polytope with vertices given by v and the midpoints of the edges. Let $v(x)$ be the vertex in whose region x falls. If x falls on the boundary of two vertex regions or at the center of mass, we assign $v(x)$ arbitrarily. Let $\varphi(x)$ be the face opposite to vertex $v(x)$, and $\eta(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ which contains x. Let $d(v(x), \eta(v(x), x))$ be the Euclidean distance from $v(x)$ to $\eta(v(x), x)$. For $r \in [1, \infty)$, let $\eta_r(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ such that

$$d(v(x), \eta_r(v(x), x)) = r\, d(v(x), \eta(v(x), x)) \quad \text{and}$$
$$d(\eta(v(x), x), \eta_r(v(x), x)) < d(v(x), \eta_r(v(x), x)).$$

Let $\mathfrak{S}_r(x)$ be the polytope similar to and with the same orientation as $\mathfrak{S}$ having $v(x)$ as a vertex and $\eta_r(v(x), x)$ as the opposite face. Then the proportional-edge proximity region $N_{PE}(x, r) := \mathfrak{S}_r(x) \cap \mathfrak{S}(\mathcal{Y}_{d+1})$. Furthermore, let $\zeta_i(x)$ be the hyperplane such that $\zeta_i(x) \cap \mathfrak{S}(\mathcal{Y}_{d+1}) \neq \emptyset$ and $r\, d(\mathsf{y}_i, \zeta_i(x)) = d(\mathsf{y}_i, \eta(\mathsf{y}_i, x))$ for $i = 1, 2, \ldots, d+1$. Then $\Gamma_1(x, r) \cap R(\mathsf{y}_i) = \{z \in R(\mathsf{y}_i) : d(\mathsf{y}_i, \eta(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \zeta_i(x))\}$, for $i = 1, 2, 3$. Hence $\Gamma_1(x, r) = \bigcup_{j=1}^{d+1} (\Gamma_1(x, r) \cap R(\mathsf{y}_i))$. Notice that $r \geq 1$ implies $x \in N_{PE}(x, r)$ and $x \in \Gamma_1(x, r)$.

Theorem 4.1 generalizes, so that any simplex $\mathfrak{S}$ in $\mathbb{R}^d$ can be transformed into a regular polytope (with edges being equal in length and faces being equal in volume) preserving uniformity. Delaunay triangulation becomes Delaunay tessellation in $\mathbb{R}^d$, provided no more than $d+1$ points being cospherical (lying on the boundary of the same sphere). In particular, with $d = 3$, the general simplex is a tetrahedron (4 vertices, 4 triangular faces and 6 edges), which can be mapped into a regular tetrahedron (4 faces are equilateral triangles) with vertices $(0, 0, 0)$ $(1, 0, 0)$ $(1/2, \sqrt{3}/2, 0)$, $(1/2, \sqrt{3}/4, \sqrt{3}/2)$.

Asymptotic normality of the edge density holds for $d > 2$ in both reflexivity and underlying graph cases.

6. Discussion and conclusions

We consider two families of random graphs based on proximity catch digraphs (PCDs) introduced recently by Ceyhan and Priebe [8]. PCDs are vertex-random graphs in the classification of Beer et al. [2] and have applications in pattern recognition and spatial data analysis. A PCD is a random digraph which is defined using proximity regions based on the relative positions of the points from various classes. Different PCDs result from different definitions of the proximity region associated with each data point. We consider the following two types of graphs associated with a given digraph: (i) underlying graph, which is obtained when any arc is replaced with an edge without allowing multi-edges [11], (ii) reflexivity graph, which is introduced in this article and is obtained when only symmetric or reflexive arcs are replaced with edges. We study the graph invariant called edge density for these graphs.

We derive the asymptotic distribution of the edge density of the graphs associated with (parameterized) proportional-edge proximity catch digraphs (PE-PCDs). In particular, we consider the reflexivity and underlying graphs associated with the PE-PCD; and derive the asymptotic distribution of their edge density using the central limit theory of U-statistics [15]. Given two classes, $\mathcal{X}$ and $\mathcal{Y}$, of points in $\mathbb{R}^2$, PE-PCDs are defined with $\mathcal{X}$ points being the vertices, and arcs being based on the relative position of $\mathcal{X}$ points with respect to Delaunay triangulation of $\mathcal{Y}$ points. In particular, PE-PCDs are defined for $\mathcal{X}$ points that lie in the convex hull of $\mathcal{Y}$ points. Hence we compute the asymptotic means and variances of the limiting normal distribution of the reflexivity and underlying graphs for uniformly distributed $\mathcal{X}$ points in the convex hull of $\mathcal{Y}$ points based on detailed geometric calculations. We also compare the asymptotic distributions of the edge densities of these graphs and that of the arc density of the PCDs. We observe that the rates of convergence to normality are sharper for the edge density of the reflexivity and underlying graphs compared to that of the arc density of the PE-PCDs. Although the calculations of the asymptotic means and variances of the arc density of the PE-PCDs and the edge density of the associated graphs are extremely tedious, the PE-PCDs have appealing properties in theory and practice. The PE-PCD (resp. the associated graphs) restricted to a Delaunay triangle forms a subdigraph (resp. subgraphs) which is not connected to the subdigraph (resp. subgraphs) restricted to any other Delaunay triangle, since the arcs (resp. the edges) of the PCDs (resp. the associated graphs)

do not cross the boundaries of the triangles by construction. Hence in our calculations, we can consider each triangle separately, rather than the entire convex hull of $\mathcal{Y}$ points. Furthermore, for uniform data in one triangle, the arc density of the PCD and the edge density of the associated reflexivity and underlying graphs do not depend on the geometry of the triangle; i.e., they are *geometry invariant* (see Theorem 4.1). Therefore, it suffices to compute the asymptotic mean and variance once for uniform data in the standard equilateral triangle (in which we can also exploit the inherent symmetry) and we can then extend the results to the multiple triangle case. We consider various extensions of the edge density to the multiple triangle case, and determine that one of the extensions has a sharper rate of convergence to normality (see Section 5).

Ceyhan et al. [10] showed that the (relative) arc density of the PE-PCDs is a U-statistic, and here it is also shown that the edge densities of the reflexivity and underlying graphs associated with the PE-PCD are also U-statistics. U-statistics in general are also closely connected to martingales and reversed martingales (see, e.g., Jurecková et al. [19]), but such connections for the arc or edge densities are not pursued here. Ceyhan et al. [10] applied the arc density of PE-PCDs for testing bivariate spatial patterns; in a similar fashion, edge densities of the associated graphs can be employed for the same purpose as well. For example, in a two-class setting, $\mathcal{X}$ and $\mathcal{Y}$ points might represent trees or plants from two different species, where former being much more abundant than the latter. So by construction, species $\mathcal{X}$ points represent the vertices of the PCD, and species $\mathcal{Y}$ points are used to construct the Delaunay triangulation which is auxiliary to define the proximity regions around species $\mathcal{X}$ points, thereby used to determine the presence or absence of an arc between any two $\mathcal{X}$ points. The arc density of the constructed PCD or the edge density of the associated graphs can be employed to test for spatial interaction between species $\mathcal{X}$ and $\mathcal{Y}$ (mostly in the direction of dependence of $\mathcal{X}$ on $\mathcal{Y}$). More specifically, the arc/edge densities can be employed for testing the complete spatial randomness (CSR) of two or more classes of points against the segregation or association of the points from the classes. *CSR* is roughly defined as the lack of spatial interaction between the points in a given study area. In particular, the null hypothesis can be assumed to be CSR of $\mathcal{X}$ points, i.e., the uniformness of $\mathcal{X}$ points in the convex hull of $\mathcal{Y}$ points. *Segregation* is the pattern in which points of one class tend to cluster together, i.e., form one-class clumps. On the other hand, *association* is the pattern in which the points of one class tend to occur more frequently around points from the other class. Under the segregation alternative, $\mathcal{X}$ points will tend to be further away from $\mathcal{Y}$ points and under the association alternative $\mathcal{X}$ points will tend to cluster around the $\mathcal{Y}$ points. Such patterns can be detected by the test statistics based on the arc/edge densities, since under segregation we expect them to be smaller, and under association they tend to be larger. Moreover, the reflexivity and underlying graphs can also be used in pattern classification as outlined in [24]. These areas are topics of ongoing research.

Having the explicit forms of the parameters of the asymptotic normal distribution of the edge density of the reflexivity and underlying graphs for uniform data allows one to perform testing uniformity of data points in the Euclidean plane. Furthermore, having the covariance structure on edges allows one to generate random graphs/digraphs that exceed the traditional random graphs/digraphs. Classical random graphs/digraphs such as Erdős–Rényi graphs/digraphs fail to approximate real life networks (such as social networks, power networks, the internet, or email traffic etc.), due to major limitations of the classical modeling. In particular, in Erdős–Rényi graphs/digraphs the edges/arcs are assumed to be iid. However, in practice, there usually is dependence between edges/arcs, at least for the ones in close vicinity (e.g., for edges/arcs sharing a vertex). Such dependence can be simulated using the parameterized PCDs or reflexivity and underlying graphs described in this article. The classical graph modeling of networks also fails in presence of clustering, triadic closure, and hubs which are commonly seen in real-life networks. As a result, alternative models such as Watts–Strogatz model [28] which aims to model the small-world properties, and the Barabási–Albert model [1], which generates scale-free networks via preferential connections, are proposed. By carefully designing the PCDs, one can replicate or simulate some of the known aspects of a network; in particular, clustering and local dependence between the arcs/edges can be conveniently accounted for. For example, in a social network, individuals may constitute vertices of the graph or the digraph, and edges or arcs between them might be inserted using some type of binary relation based on certain characteristics. However, spatially or socially "closer" individuals might be more likely to satisfy the criterion to have an edge or arc between them, but these edges or arcs are more likely to be dependent

rather than independent. If the level of connection between individuals or nodes (i.e., arc or edge probability) of the network digraph or graph is known (or approximately estimated) and most of the dependence between any two pairs of edges/arcs occurs when one vertex is shared, the corresponding network can be simulated (at least approximately) by using a PE-PCD or a related graph. Here, one can even simulate the disconnected or weakly connected clusters as components. Each disconnected component might be represented by the points generated in one triangle; and weakly connected components (i.e., components connected by few edges/arcs compared to the number of edges/arcs in the components) might be simulated by points generated in as many triangles as the number of components, with randomly inserting edges/arcs that cross one triangle to another (i.e., edges/arcs between vertices residing in different possibly neighboring triangles) at the desired numbers. These aspects are among the topics of prospective research. Finally, the methodology described here is also applicable to PCDs in higher dimensions.

Acknowledgments

I would like to thank an anonymous associate editor and referee, whose constructive comments and suggestions greatly improved the presentation and flow of the paper. Most of the Monte Carlo simulations presented in this article were executed at Koç University High Performance Computing Laboratory. This research was supported by the European Commission under the Marie Curie International Outgoing Fellowship Programme via Project #329370 titled PRinHDD.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.stamet.2016.07.003.

References

[1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[2] E. Beer, J.A. Fill, S. Janson, E.R. Scheinerman, On vertex, edge, and vertex-edge random graphs, Electron. J. Combin. 18 (2011) #P110.
[3] M. Boullé, Nonparametric edge density estimation in large graphs. Technical report, No FT/RD/TECH/11/02/13, France Telecom R&D, 2011.
[4] H. Callaert, P. Janssen, The Berry–Esseen theorem for U-statistics, Ann. Statist. 6 (1978) 417–421.
[5] A. Cannon, L. Cowen, Approximation algorithms for the class cover problem, in: Proceedings of the 6th International Symposium on Artificial Intelligence and Mathematics, January 5–7, 2000, Fort Lauderdale, Florida, 2000.
[6] E. Ceyhan, Extension of one-dimensional proximity regions to higher dimensions, Comput. Geom. 43 (9) (2010) 721–748.
[7] E. Ceyhan, Distribution of relative edge density of the underlying graphs based on a random digraph family. Technical Report # KU-EC-13-3, Koç University, Istanbul, Turkey, 2014, arXiv:1002.2957v2 [math.PR].
[8] E. Ceyhan, C.E. Priebe, The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association, Statist. Probab. Lett. 73 (2005) 37–50.
[9] E. Ceyhan, C.E. Priebe, D.J. Marchette, A new family of random graphs for testing spatial segregation, Canad. J. Statist. 35 (1) (2007) 27–50.
[10] E. Ceyhan, C.E. Priebe, J.C. Wierman, Relative density of the random r-factor proximity catch digraphs for testing spatial patterns of segregation and association, Comput. Statist. Data Anal. 50 (8) (2006) 1925–1964.
[11] G. Chartrand, L. Lesniak, P. Zhang, Graphs & Digraphs, fifth ed., Chapman and Hall/CRC, Boca Raton, Florida, 2010.
[12] R.K. Darst, D.R. Reichman, P. Ronhovde, Z. Nussinov, An edge density definition of overlapping and weighted graph communities. Technical Report, 2013, e-print arXiv:1301.3120.
[13] J. DeVinney, C.E. Priebe, A new family of proximity graphs: Class cover catch digraphs, Discrete Appl. Math. 154 (14) (2006) 1975–1982.
[14] B. Grünbaum, The edge-density of 4-critical planar graphs, Combinatorica 8 (1) (1988) 137–139.
[15] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Stat. 19 (1948) 293–325.
[16] W. Hoeffding, The strong law of large numbers for U-statistics. Institute of Statistics Mimeo Series #326, Raleigh, NC, 1961.
[17] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58 (1963) 13–30.
[18] J.W. Jaromczyk, G.T. Toussaint, Relative neighborhood graphs and their relatives, Proc. IEEE 80 (1992) 1502–1517.
[19] J. Jurecková, P.K. Sen, J. Picek, Methodology in Robust and Nonparametric Statistics, CRC Press, Boca Raton, Florida, 2012.
[20] P. Keevash, B. Sudakov, Local density in graphs with forbidden subgraphs, Combin. Probab. Comput. 12 (2003) 139–153.
[21] T.S. Michael, Sphere of influence graphs: edge density and clique size, Math. Comput. Modelling 20 (4) (1994) 19–24.
[22] M. Penrose, Random Geometric Graphs, in: Oxford Studies in Probability, vol. 5, Oxford University Press, Oxford, UK, 2003.
[23] C.E. Priebe, J.G. DeVinney, D.J. Marchette, On the distribution of the domination number of random class cover catch digraphs, Statist. Probab. Lett. 55 (2001) 239–246.

[24] C.E. Priebe, D.J. Marchette, J. DeVinney, D. Socolinsky, Classification using class cover catch digraphs, J. Classification 20 (1) (2003) 3–23.
[25] C.E. Priebe, J.L. Solka, D.J. Marchette, B.T. Clark, Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays, Comput. Statist. Data Anal. 43 (4) (2003) 621–632.
[26] A.A. Razborov, On the minimal density of triangles in graphs, Combin. Probab. Comput. 17 (4) (2008) 603–618.
[27] G.T. Toussaint, The relative neighborhood graph of a finite planar set, Pattern Recognit. 12 (4) (1980) 261–268.
[28] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.