Statistics & Probability Letters 55 (2001) 239–246

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

On the distribution of the domination number for random class cover catch digraphs ☆

Carey E. Priebe *, Jason G. DeVinney, David J. Marchette

Department of Mathematical Sciences, Whiting School of Engineering, Johns Hopkins University, 3400 N, Charles Street, 104 Whitehead Hall, Baltimore, MD 21218-2682, USA

Received October 2000; received in revised form March 2001

Abstract

In this article we initiate the study of *class cover catch digraphs*, a special case of intersection digraphs motivated by applications in machine learning and statistical pattern recognition. Our main result is the exact distribution of the domination number for a data-driven model of random interval catch digraphs. © 2001 Elsevier Science B.V. All rights reserved

Keywords: Random graph; Intersection digraph; Domination number; Distribution

1. Class cover catch digraphs

Let $\mathscr{X} = \{X_1, \ldots, X_n\}$, $\mathscr{Y} = \{Y_1, \ldots, Y_m\}$ be two classes of $\mathbb{R}^q$-valued random variables with joint distribution $F_{X,Y}$. Let $d(\cdot,\cdot): \mathbb{R}^q \times \mathbb{R}^q \to [0,\infty)$ be any distance function. The *class cover problem* for a distinguished target class (say, $\mathscr{X}$; note that the problem is asymmetric in target class) is to find a collection of open balls $B_i := B(c_i, r_i) := \{x: d(x, c_i) < r_i\}$ such that (i) $\mathscr{X} \subset \cup_i B_i$ and (ii) $\mathscr{Y} \cap \cup_i B_i = \emptyset$. A collection of balls satisfying (i) and (ii) is termed a *class cover*. Condition (i) defines a *proper cover of class* $\mathscr{X}$ while (ii) defines a *pure cover with respect to class* $\mathscr{Y}$. A *constrained* class cover requires the balls to be centered at target class observations; $c_i \in \mathscr{X}\ \forall i$. A *homogeneous* class cover requires the ball radii to be the same for all balls; $r_i = r\ \forall i$. This article investigates minimum cardinality constrained heterogeneous class covers—class covers which satisfy (i) and (ii) with the fewest balls possible. This class cover problem, a generalization of the set cover problem (see, e.g., Garfinkel and Nemhauser, 1972) motivated by applications in machine learning and statistical pattern recognition wherein the methodology requires the selection of a (small) set

☆ Supported by the Defense Advanced Research Projects Agency as administered by the Air Force Office of Scientific Research under contract DOD F49620-99-1-0213 and by Office of Naval Research Grant N00014-95-1-0777.

* Corresponding author. Tel.: +1-410-516-7002; fax: +1-410-516-7459.
E-mail address: cep@jhu.edu (C.E. Priebe).

PII: S0167-7152(01)00129-8

of representative exemplars from $\mathcal{X}$, has been previously considered in Cannon and Cowen (2001) and Priebe and Marchette (2000).

Definition. The random *class cover catch digraph G for $\mathcal{X}$, $\mathcal{Y}$* corresponding to the constrained heterogeneous class cover problem for target class $\mathcal{X}$ is the digraph of order n with vertex set $\mathcal{X}$ and a directed edge from X_i to X_j if and only if $X_j \in B(X_i, \min_{Y\in\mathcal{Y}} d(X_i, Y))$. That is, there is an edge from the ith vertex to the jth vertex if and only if there exists an open ball centered at X_i which is "pure", or contains no elements of class $\mathcal{Y}$, and simultaneously contains, or "catches", point X_j. We write $V(G)=\{v_i,\ldots,v_n\}$ with each v_i corresponding to X_i; thus $v_iv_j \in E(G)$ if and only if $X_j \in B_i$. We call such a digraph a random *$\mathcal{C}_{n,m}$-graph*. (Some authors e.g., Chartrand and Lesniak, 1996 use the term *pseudodigraphs* since $\mathcal{C}_{n,m}$-graphs permit loops.)

Our $\mathcal{C}_{n,m}$-graphs can be seen to be a special case of both the intersection digraphs of Sen et al. (1989) and the covering sets or transversals of Tuza (1994). In addition, significant similarities are apparent between $\mathcal{C}_{n,m}$-graphs and *proximity* and *neighborhood* graphs (Jaromczyk and Toussaint, 1992), *sphere-of-influence* graphs (see, e.g., McKee and McMorris, 1999), and the *sphere-of-attraction* graphs presented in McMorris and Wang (2000). Our model is a vertex-random graph model (Karoński et al., 1999) and is not one of the standard models (see, e.g., Bollobas, 1985 or Janson et al., 2000). Rather, the randomness in our $\mathcal{C}_{n,m}$ model resides in the vertices $\mathcal{X}$ and the existence of an edge v_iv_j is a deterministic function of the random variable X_j and the random set B_i. Thus our random graph model is *data-driven*—a function of the joint distribution $F_{X,Y}$.

2. Domination number for random $\mathcal{C}_{n,m}$-graphs

Definition. For a pseudodigraph $G=(V,E)$, a *dominating set* is a collection of vertices $D \subset V$ such that for each $v_i \in V$ either (1) $v_i \in D$ or (2) there is an edge from some $v_j \in D$ to v_i ($v_jv_i \in E$). That is, $N[D]=V$, where $N[D]$ denotes the closed neighborhood of $D \subset V$. See, for instance, Bollobas (1998), Chartrand and Lesniak (1996), Haynes et al. (1998a, b). A *minimum dominating set* is a dominating set D for which the cardinality $|D|$ is minimum. The cardinality of a minimum dominating set is the *domination number*, denoted $\gamma(G)$. There is always a dominating set of cardinality $|V|$; namely, $D=V$. Thus $\gamma(G) \leqslant n$.

Definition. For a pseudodigraph $G=(V,E)$, a *total dominating set* is $D \subset V$ such that for each $v_i \in V$ there exists $v_j \in D$ such that $v_jv_i \in E$. That is, $N(D)=V$, where $N(D)$ denotes the open neighborhood of $D \subset V$. Note that if there is a vertex $v_i \in V$ with no in-edge, from itself or from any other $v_j \in V$, then no total dominating set exists for G. The invariant $\gamma_t(G)$, the *total domination number*, represents the cardinality of a minimum total dominating set. If no total dominating set exists, we say $\gamma_t(G)=\infty$.

If the joint distribution $F_{X,Y}$ satisfies the assumption that $P[X=Y]=0$—for instance, if X and Y are independent and the class-conditional probability density functions f_X, f_Y exist—then $\gamma_t(G)=\gamma(G)$ almost surely (a.s.) for $G \in \mathcal{C}_{n,m}$. Since $X_i \in B_i$ unless the open ball $B_i=\emptyset$ or, equivalently, unless $r_i{:=}\min_{Y\in\mathcal{Y}} d(X_i, Y)=0$, it follows that every v_i has an in-edge from itself ($v_iv_i \in E$) if and only if it is not the case that there are a class $\mathcal{X}$ observation and a class $\mathcal{Y}$ observation which are coincident. Furthermore, $v_iv_i \notin E \Rightarrow v_jv_i \notin E$ for any $v_j \in V$. Thus the existence of a vertex v_i such that $v_iv_i \notin E$ implies $\gamma_t(G)=\infty$, and represents the only case for which $\gamma_t(G)$ and $\gamma(G)$ differ.

For the purposes of this article we consider "nice" distributions—joint distributions $F_{X,Y}$ for which $\gamma_t(G)<\infty$ a.s. That is, we consider $\mathcal{F}(\mathbb{R}^q)$-random $\mathcal{C}_{n,m}$-graphs, defined as random $\mathcal{C}_{n,m}$-graphs for which $F_{X,Y} \in \mathcal{F}(\mathbb{R}^q)$, where

$$\mathcal{F}(\mathbb{R}^q){:=}\{F_{X,Y} \text{ on } \mathbb{R}^q \text{ such that } P[X=Y]=0\}.$$

The following theorem follows immediately:

Theorem 1. *Let G be an $\mathscr{F}(\mathbb{R}^q)$-random $\mathscr{C}_{n,m}$-graph. Then $\gamma_t(G)=\gamma(G)$ a.s.*

In light of the result given in Theorem 1, we henceforth focus our attention on the random variable $\gamma(G)$ for $\mathscr{F}(\mathbb{R}^q)$-random $\mathscr{C}_{n,m}$-graphs. We will assume throughout that n and m are positive integers and use the notation $\gamma(G;n,m)$ for $\gamma(G)$ to make explicit the dependence on the sample sizes. The trivial result $1 \leqslant \gamma(G;n,m) \leqslant n$ is immediately available.

3. $\mathscr{F}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graphs

Consider now class cover catch digraphs in one dimension—the case $q=1$. Class cover catch digraphs on $\mathbb{R}$ are a special case of *interval catch digraphs* (Sen et al., 1989; Prisner, 1989; Prisner, 1994).

Consider the collection of $m+1$ intervals based on the order statistics

$$-\infty =: Y_{(0:m)} < Y_{(1:m)} \leqslant Y_{(2:m)} \leqslant \cdots \leqslant Y_{(m:m)} < Y_{(m+1:m)} := +\infty;$$

$I_j := (Y_{(j-1:m)}, Y_{(j:m)})$ for $j=1,\ldots,m+1$. Let $\mathscr{X}_j = I_j \cap \mathscr{X}$ and $\mathscr{Y}_j = \{Y_{(j-1:m)}, Y_{(j:m)}\}$. This gives us $m+1$ disconnected subgraphs G_j, each of which may be null or may itself be disconnected into at most two components. Define $N_j := |\mathscr{X}_j|$, and let $\gamma_j(N_j)$ denote the cardinality of a minimum dominating set for the random $\mathscr{C}_{n,m}$-graph induced by $\mathscr{X}_j$, $\mathscr{Y}_j$. Then $\gamma(G;n,m)=\sum_{j=1}^{m+1}\gamma_j(N_j)$. Thus the study of $\gamma(G;n,m)$ is carried out via the investigation of the simpler random variables $\gamma_j(N_j)$.

Lemma 1. *For $j=1$, $m+1$ we have $\gamma_j(N_j)=1\{N_j>0\}$. ($1\{\cdot\}$ is the indicator function.)*

Proof. Clearly $\gamma_j(0)=0$. Consider $j=1$ and the case $N_1 \geqslant 1$. Define $B_1 := B(\min(\mathscr{X}_1), Y_{(1:m)} - \min(\mathscr{X}_1))$, the largest pure open ball centered at the leftmost observation in I_1. Then $\mathscr{X}_1 \subset B_1$ and $\mathscr{Y}_1 \cap B_1 = \emptyset$, and hence $\gamma_1(N_1)=1$. The case $j=m+1$ follows similarly. □

For $j=2,\ldots,m$ we now show that, for $N_j \geqslant 1$, $\gamma_j(N_j)$ takes values in $\{1,2\}$ with distribution-dependent probabilities $\{\kappa_j(N_j), 1-\kappa_j(N_j)\}$, respectively, where

$$\kappa_j(N_j) := P[\mathscr{X}_j \cap I_j^* \neq \emptyset]$$

with

$$I_j^* := \left(\frac{\max(\mathscr{X}_j)+Y_{(j-1:m)}}{2}, \frac{\min(\mathscr{X}_j)+Y_{(j:m)}}{2}\right) \subset I_j.$$

Lemma 2. *For $j=2,\ldots,m$, $\gamma_j(N_j) =^{\mathrm{d}} 1 + \mathrm{Bernoulli}(1-\kappa_j(N_j))$ for $N_j \geqslant 1$.*

Proof. Again, $\gamma_j(0)=0$. If $N_j=1$ ($\mathscr{X}_j=\{X\}$) then there exists an $\varepsilon>0$ such that the ball $B(X,\varepsilon)$ suffices to demonstrate that $\gamma_j(1)=1$, and $X \in I_j^*$ so $\kappa_j(1)=1$ as desired. Suppose now that $N_j \geqslant 2$. Let $X_j^- := \max\{X \in \mathscr{X}_j\colon X \leqslant (Y_{(j-1:m)}+Y_{(j:m)})/2\}$ and $X_j^+ := \min\{X \in \mathscr{X}_j\colon X \geqslant Y_{(j-1:m)}+Y_{(j:m)}/2\}$. ($X_j^- < X_j^+$ a.s. if both exist.) Let $B_j^- := B(X_j^-, X_j^- - Y_{(j-1:m)})$ and $B_j^+ := B(X_j^+, Y_{(j:m)} - X_j^+)$. ($B_j^-$ (B_j^+) $=\emptyset$ if X_j^- (X_j^+) does not exist.) Since $\mathscr{X}_j \subset (B_j^- \cup B_j^+)$ and $\mathscr{Y}_j \cap (B_j^- \cup B_j^+) = \emptyset$, it follows that $\gamma_j(N_j) \in \{1,2\}$. Finally, observe that $\gamma_j(N_j)=1 \Leftrightarrow$ there exists $X \in \mathscr{X}_j$ such that (i) $X - \min(X_j) < Y_{(j:m)} - X$ and (ii) $\max(X_j) - X < X - Y_{(j-1:m)}$, and (i) and (ii) hold if and only if there exists $X \in I_j^*$. □

Clearly $P[\mathscr{X}_j \cap I_j^* \neq \emptyset]$ depends on the conditional distribution $F_{X|Y}$ on the interval I_j; if we know this distribution, we can calculate κ_j.

As an immediate consequence of the preceding two Lemmas the upper bound of n for $\gamma(G;n,m)$ can be tightened for $\mathscr{F}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graphs.

Theorem 2. *Let G be an $\mathscr{F}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graph. Then* $1 \leqslant \gamma(G;n,m) \leqslant \min(n,2m)$.

3.1. Main result

The following lemma provides the exact result for the distribution of the cardinality of a minimum dominating set for a simple random $\mathscr{C}_{n,2}$-graph. Our main result will employ this lemma.

Lemma 3. *Let $-\infty < a < b < +\infty$. Let $\mathscr{X} = \{X_1,\dots,X_n\}$ with $X_i \sim^{\text{i.i.d.}} \text{Uniform}(a,b)$. Let $\mathscr{Y} = \{a,b\}$. Let G be the random $\mathscr{C}_{n,2}$-graph for $\mathscr{X}$, $\mathscr{Y}$. Then $\gamma(G;n,2) =^{\text{d}} 1 + \text{Bernoulli}(1-\kappa(n))$ with*

$$\kappa(n) := P[\gamma(G;n,2)=1] = \tfrac{5}{9} + \tfrac{4}{9}4^{-(n-1)}$$

for $n \geqslant 1$. (For future reference, we note that $\kappa(1)=1$, $\kappa(2)=\frac{6}{9}$, and $\kappa(n) \searrow \frac{5}{9}$ as $n \nearrow \infty$.)

Proof. Let $I^* := ((\max(\mathscr{X})+a)/2, (\min(\mathscr{X})+b)/2) \subset (a,b)$. By Lemma 2, $\gamma(G;n,2)=1 \Leftrightarrow \mathscr{X} \cap I^* \neq \emptyset$ and it suffices to show that $P[\mathscr{X} \cap I^* \neq \emptyset] = \frac{5}{9} + \frac{4}{9}\cdot 4^{-(n-1)}$. For this, note first that since the X_i are independent and identically distributed uniform over (a,b), the random variable $1\{\mathscr{X} \cap I^* = \emptyset\}$ is independent of a, b, and the size of the interval $b-a$. Thus, to simplify our calculations, we set $a=0$ and $b=1$. To further simplify our calculations we calculate $P[\mathscr{X} \cap I^* \neq \emptyset]$ by calculating $1 - P[\mathscr{X} \cap I^* = \emptyset]$.

We condition on $\min(\mathscr{X})$ and $\max(\mathscr{X})$, writing $\min(\mathscr{X}) = x_1$ and $\max(\mathscr{X}) = x_n$, to obtain

$$P[\mathscr{X} \cap I^* = \emptyset] = \int_0^{1/2} \int_{\max\{(x_1+1)/2, 2x_1\}}^1 f_{x_1,x_n}(x_1,x_n) g(x_1,x_n)\, \mathrm{d}x_n\, \mathrm{d}x_1,$$

where $f_{x_1,x_n}(x_1,x_n) = n(n-1)(x_n-x_1)^{n-2}$ is the joint probability density function of $\min(\mathscr{X})$ and $\max(\mathscr{X})$ (see, e.g., David, 1981) and for $y > \max\{(x+1)/2, 2x\}$

$$g(x,y) := P[\mathscr{X} \cap I^* = \emptyset \mid \min(\mathscr{X})=x, \max(\mathscr{X})=y] = \left(1 - \left(\frac{1+x-y}{2(y-x)}\right)\right)^{n-2}.$$

Thus we have

$$\begin{aligned} P[\mathscr{X} \cap I^* = \emptyset] &= \int_0^{1/3} \int_{(1+x_1)/2}^1 f_{x_1,x_n}(x_1,x_n) g(x_1,x_n)\, \mathrm{d}x_n\, \mathrm{d}x_1 \\ &\quad + \int_{1/3}^{1/2} \int_{2x_1}^1 f_{x_1,x_n}(x_1,x_n) g(x_1,x_n)\, \mathrm{d}x_n\, \mathrm{d}x_1 \\ &= \frac{1}{9}(4 - 2^{2-n} - 2^{3-2n}) + \frac{1}{9}(2^{2-n} - 2^{3-2n}) \\ &= \frac{4}{9} - \frac{4}{9}\cdot 4^{-(n-1)}. \end{aligned}$$

The desired result follows immediately. □

Lemma 3 allows the derivation of the distribution of $\gamma(G;n,m)$ for a particular class of random $\mathscr{C}_{n,m}$-graphs on $\mathbb{R}$. Define $\mathscr{U}(\mathbb{R}) \subset \mathscr{F}(\mathbb{R})$ as

$$\mathscr{U}(\mathbb{R}) := \{F_{X,Y}\colon X \text{ and } Y \text{ are independent}, X_i \sim^{\text{i.i.d}} F_X,\ Y_j \sim^{\text{i.i.d}} F_Y, \text{ and}$$
$$F_X = F_Y = \text{Uniform } (I) \text{ where } I=(a,b) \subset \mathbb{R} \text{ with } -\infty < a < b < +\infty\}.$$

The following theorem gives the exact distribution of the cardinality of minimum total dominating sets for $\mathscr{U}(\mathbb{R})$—random class cover catch digraphs.

Let Z_m denote the set of non-negative integers less than m; $Z_m := \{0,\ldots,m-1\}$. Define

$$\Delta_{z,b}^{S} := \left\{ (z_1,\ldots,z_b) \colon \sum_{i=1}^{b} z_i = z; z_i \in S \ \forall i \right\}.$$

Theorem 3. *Let G be a $\mathscr{U}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graph. Then the probability mass function for the random variable $\gamma(G;n,m)$ is given by*

$$P[\gamma(G;n,m)=d] = \frac{n!m!}{(n+m)!} \sum_{\vec{n}\in\Delta_{n,m+1}^{Z_{n+1}}} \sum_{\vec{d}\in\Delta_{d,m+1}^{Z_3}} \alpha(d_1,n_1)\cdot\alpha(d_{m+1},n_{m+1}) \prod_{j=2}^{m} \beta(d_j,n_j),$$

where

$$\alpha(d,n) = \max(1\{n=d=0\}, 1\{n \geqslant d=1\})$$

and

$$\beta(d,n) = \max(1\{n=d=0\}, 1\{n \geqslant d \geqslant 1\}) \cdot \kappa(n)^{1\{d=1\}} \cdot (1-\kappa(n))^{1\{d=2\}}.$$

Proof. For $\gamma(G;n,m) = \sum_{j=1}^{m+1} \gamma_j(n_j) = d$ we must have $\gamma_1(n_1) = d_1$ and $\cdots$ and $\gamma_{m+1}(n_{m+1}) = d_{m+1}$ for some $\vec{d} = (d_1,\ldots,d_{m+1})$ such that $\sum_{j=1}^{m+1} d_j = d$ and some $\vec{n} = (n_1,\ldots,n_{m+1})$ such that $\sum_{j=1}^{m+1} n_j = n$. $\Delta_{n,m+1}^{Z_{n+1}}$ is precisely the collection of $\vec{n}$ which can occur and, since the individual d_j can take values only in $\{0,1,2\}$, $\Delta_{d,m+1}^{Z_3}$ is precisely the collection of $\vec{d}$ which can occur. Therefore we have

$$P[\gamma(G;n,m)=d] = \sum_{\vec{n}\in\Delta_{n,m+1}^{Z_{n+1}}} \sum_{\vec{d}\in\Delta_{d,m+1}^{Z_3}} P[\vec{n}] \prod_{j=1}^{m+1} P[\gamma_j(n_j)=d_j|\vec{n}]$$

$$= \sum_{\vec{n}\in\Delta_{n,m+1}^{Z_{n+1}}} \sum_{\vec{d}\in\Delta_{d,m+1}^{Z_3}} P[\vec{n}] \prod_{j\in\{1,m+1\}} P[\gamma_j(n_j)=d_j|\vec{n}] \prod_{j=2}^{m} P[\gamma_j(n_j)=d_j|\vec{n}],$$

where we have used the conditional pairwise independence of the γ_j. The form of the final expression is due to the fact that we need to treat the end intervals I_1 and I_{m+1} separately. Certain pairs (n_j,d_j) are incompatible, such as $n_j = 0$ and $d_j > 0$; the indicator functions in the statement of Theorem 3 eliminate incompatible pairs from the summation. For the end intervals I_1 and I_{m+1}, the α terms yield a value of unity if the (n_j,d_j) pair is compatible. The β terms are derived from compatibility considerations and the result of Lemma 3. The desired result is obtained by noting that each $\vec{n}\in\Delta_{n,m+1}^{Z_{n+1}}$ has probability $1/\binom{n+m}{n}$ of occurring. □

While the expected value of $\gamma(G;n,m)$ can be obtained from Theorem 3, a more straightforward derivation is provided.

Theorem 4. *Let G be a $\mathscr{U}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graph. Then*

$$E[\gamma(G;n,m)] = \frac{2n}{n+m} + \frac{n!m(m-1)}{(n+m)!} \sum_{i=1}^{n} \frac{(n+m-i-1)!}{(n-i)!}(2-\kappa(i)),$$

where $\kappa(i)$ is given by Lemma 3.

Proof. The expected value of $\gamma(G;n,m)$ for an $\mathscr{F}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graph G is given by

$$E[\gamma(G;n,m)] = P[X_{(1:n)} < Y_{(1:m)}] + \sum_{j=2}^{m}\sum_{i=1}^{n} P[N_j = i]E[\gamma_j(i)] + P[X_{(n:n)} > Y_{(m:m)}]$$

$$= \frac{2n}{n+m} + (m-1)\sum_{i=1}^{n} P[N_2 = i]E[\gamma_2(i)].$$

For $j = 2,\dots,m$ we have $\gamma_j(N_j) \sim 1 + \text{Bernoulli}(1-\kappa(N_j))$ from Lemma 3 and thus $E[\gamma_j(i)] = (2-\kappa_j(i))$, and $P[N_j = i] = n!m(n+m-i-1)!/(n+m)!(n-i)!$. For $j = 1, m+1$ we have $\gamma_j(N_j) = 1\{N_j > 0\}$, and $P[N_1 > 0] = P[N_{m+1} > 0] = n/n+m$. □

Corollary 1. *$E[\gamma(G;n,n)] \to \infty$ as $n \to \infty$.*

Proof. From Theorem 4 we have

$$E[\gamma(G;n,n)] = 1 + \frac{n!n(n-1)}{(2n)!}\sum_{i=1}^{n}\frac{(2n-i-1)!}{(n-i)!}(2-\kappa(i))$$

$$= 1 + \frac{(n!)^2(n-1)}{(2n)!}\sum_{i=1}^{n}\binom{2n-i-1}{n-i}(2-\kappa(i))$$

$$\geqslant 1 + \frac{(n!)^2(n-1)}{(2n)!}\sum_{i=1}^{n}\binom{2n-i-1}{n-i}$$

$$= 1 + \frac{(n!)^2(n-1)}{(2n)!}\frac{2^{n-1}(2n-1)!!}{n!}$$

$$= \frac{n+1}{2}. \qquad \square$$

Limiting results for $\gamma(G;n,m)$ are also available.

Theorem 5. (i) *Let $G_{n,m}$ be a $\mathscr{U}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graph. Then for n fixed and finite*, $\lim_{m\to\infty}\gamma(G_{n,m}) = n$ *a.s.* (ii) *Let $G_{n,m}$ be a $\mathscr{U}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graph. Then for m fixed and finite*, $\lim_{n\to\infty}\gamma(G_{n,m}) \stackrel{d}{=} m+1+B$, *where $B \sim$ Binomial$(m-1,\frac{4}{9})$.*

Proof. (i) is straightforward. For (ii), note that as $n \to \infty$, $N_j \to \infty\ \forall j$ a.s. Thus $\lim\gamma_1 = 1$ a.s., $\lim\gamma_{m+1} = 1$ a.s., and $\lim\gamma_j = 1 + \text{Bernoulli}(\frac{4}{9})$ a.s. for $j = 2,\dots,m$. Finally, note again that, conditional on the N_j, the γ_j are independent. □

3.2. A greedy algorithm yields $\hat{\gamma}(G) = \gamma(G)$

The random class cover catch digraph G for $\mathscr{X}$, $\mathscr{Y}$ can be described via the $n \times n$ adjacency matrix $A = A(G) = [a_{i,j}]$ where $a_{i,j} = 1\{d(X_i,X_j) < \min_{Y\in\mathscr{Y}} d(X_i,Y)\}$. For a graph $G \in \mathscr{C}_{n,m}$ with adjacency matrix A, $\gamma(G)$ is given by $\vec{1}^{\mathrm{T}}\vec{x}^*$ for a solution $\vec{x}^*$ to the linear algebra optimization problem

$$\begin{aligned} \min \quad & \vec{1}^{\mathrm{T}}\vec{x} \\ \text{s.t.} \quad & A^{\mathrm{T}}\vec{x} > \vec{0}. \\ & \text{N.B. } \vec{x} \in \{0,1\}^n. \end{aligned}$$

Here the elements x_i of the binary vector $\vec{x}$ of length n indicate whether the ith vertex is in the dominating set. (If no solution $\vec{x}^*$ exists then no dominating set exists; thus Theorem 1 says that a solution exists a.s.)

The linear algebra optimization problem presented above is in general an NP-hard optimization problem (Karp, 1972; Arora and Lund, 1997). A deterministic "greedy heuristic" approximation algorithm, similar to the algorithm discussed in Chvatal (1979) and Parekh (1991), for finding small dominating sets in graphs proceeds as follows:

Given $n \times n$ adjacency matrix $A = A(G)$ with rows $\vec{a}_{i,\cdot}$ and columns $\vec{a}_{\cdot,j}$
If trace$(A) < n$ Return $\hat{\gamma}(G) = \infty$; Else
Set $d = 0$
While $\vec{1}^{\mathrm{T}} A \vec{1} > \vec{0}$
  Set $d = d + 1$
  Set $i_d^* = \min \arg\max_i \vec{1}^{\mathrm{T}} \vec{a}_{i,\cdot}$
  Set $\vec{a}_{\cdot,j} = \vec{0}$ for j such that $\vec{a}_{i_d^*,j} = 1$
EndWhile
Return $\hat{\gamma}(G) = d$, $D = \{i_1^*, \ldots, i_d^*\}$

This greedy algorithm is guaranteed to find a dominating set if one exists. The algorithm is *not* guaranteed to find a minimum dominating set in the general setting. Thus we have $\hat{\gamma}(G) \geqslant \gamma(G)$. Running time analysis and approximation properties for similar greedy algorithms have been previously considered; see, e.g., Arora and Lund (1997), Chvatal (1979), Cannon and Cowen (2001), Hochbaum (1982), Johnson (1974), Nikoletseas and Spirakis (1994), Parekh (1991), Prisner (1994).

Theorem 6. *The greedy algorithm is optimal for finding a minimum dominating set for $\mathscr{F}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graphs. That is, $\hat{\gamma}(G) = \gamma(G)$ a.s. when G is an $\mathscr{F}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graph.*

Proof. It suffices to consider the subgraph G_j induced by the $n_j > 0$ observations on a single interval I_j. If $\gamma(G_j) = 1$, then $\max_i \vec{1}^{\mathrm{T}} \vec{a}_{i,\cdot} = n_j$. Otherwise, the first ball selected by the greedy algorithm (i^* for $d = 1$) either covers at least all observations to the left of the midpoint of I_j or covers at least all observations to the right of the midpoint of I_j. Then the second ball covers at least the remaining observations. □

(Note that a simpler, linear time algorithm is available for $\mathscr{F}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graphs, as suggested by the proof of Lemma 2.)

4. Discussion

Much of the machinery of Hilbert or Banach spaces is unnecessary in our general definition of class cover catch digraphs. For instance, an inner product is unnecessary, a semi-norm suffices as the triangle inequality is not used, the completeness of the space is not necessary, and our $\mathscr{C}_{n,m}$ model is appropriate for infinite dimensional (function-valued) random variables. Furthermore, the finiteness condition on the sample sizes n and m can be relaxed, giving rise to well-defined infinite pseudodigraphs.

However, it is the one-dimensionality of $\mathbb{R}$ that is fundamental to our main results. For example, the following generalization is instructive. Let S_1 be the sphere in $\mathbb{R}^2$, and define $\mathscr{U}(S_1)$-random $\mathscr{C}_{n,m}$-graphs in a manner analogous to the $\mathscr{U}(\mathbb{R})$-random $\mathscr{C}_{n,m}$-graphs. Then, in analogy with Theorem 5, we have the following result.

Theorem 7. (i) *Let $G_{n,m}$ be a $\mathscr{U}(S_1)$-random $\mathscr{C}_{n,m}$-graph. Then for n fixed and finite*, $\lim_{m\to\infty} \gamma(G_{n,m}) = n$ *a.s.* (ii) *Let $G_{n,m}$ be a $\mathscr{U}(S_1)$-random $\mathscr{C}_{n,m}$-graph. Then for m fixed and finite*, $\lim_{n\to\infty} \gamma(G_{n,m}) =^{\mathrm{d}} m + B$, *where $B \sim Binomial(m, \frac{4}{9})$.*

Our main result can be generalized in terms of the random models to which it applies. Consider the case for which X and Y are independent and continuous. If F_Y is known then the distribution of the location and size of the intervals I_j can be derived. Then, as noted above, if F_X is known one can calculate the κ_j.

Our main result applies to the constrained heterogeneous class cover problem in $\mathbb{R}$. The unconstrained heterogeneous problem is trivial, while the unconstrained homogeneous problem depends only on the interval sizes. Investigation of the domination number for the constrained homogeneous problem is of some interest (see Cannon and Cowen, 2001 and Priebe and Marchette, 2000).

Applications in statistical pattern recognition and machine learning demand the investigation into higher dimensional class cover catch digraphs. This investigation is continuing, with Guibas et al. (1994) providing the results of a similar investigation for sphere-of-influence graphs. Characterization of class cover catch digraphs on $\mathbb{R}^q$, as a function of q, is ongoing.

Finally, a generalization of our class cover catch digraphs is being investigated wherein the requirements for a pure and proper cover are relaxed. This generalization is of particular interest in the applications under consideration due to the need to avoid overfitting in the class cover. Specifically, given parameters $\alpha, \beta \in [0, 1]$ we require (i′) $|\mathcal{X} \cap \cup_i B_i| \geqslant n\alpha$ and (ii′) $|\mathcal{Y} \cap \cup_i B_i| \leqslant m\beta$. The former condition relaxes the requirement that all target class observations be covered, while the latter condition allows some non-target class observations to be covered. See Priebe and Marchette (2000) and, analogously, Guibas et al. (1994). The case $\alpha = 1$ and $\beta = 0$ reduces to the problem studied herein. This generalization still can be seen to be a special case of both the intersection digraphs of Sen et al. (1989) and the transversals of Tuza (1994). In this generalization, and with the possibility of coincident observations of differing class ($P[X = Y] > 0$), some vertices may have loops while others do not.

References

Arora, S., Lund, C., 1997. Hardness of approximations. PWS Publishing Company, Boston, MA.
Bollobas, B., 1985. Random Graphs. Academic Press, New York.
Bollobas, B., 1998. Modern Graph Theory. Springer, Berlin.
Cannon, A., Cowen, L.C., 2001. Approximation algorithms for the class cover problem. Ann. Math. AI. 30, to appear.
Chartrand, G., Lesniak, L., 1996. Graphs and Digraphs. Chapman & Hall, London.
Chvatal, V., 1979. A greedy heuristic for the set-covering problem. Mathematics of Operations Research 4, 233–235.
David, H.A., 1981. Order Statistics. Wiley, New York.
Garfinkel, R.S., Nemhauser, G.L., 1972. Integer Programming. Wiley, New York.
Guibas, L., Pach, J., Sharir, M., 1994. Sphere of influence graphs in higher dimensions. Colloquia Mathematica Societatis Janos Bolyai: Intuitive Geometry 63, 131–137.
Haynes, T.W., Hedetniemi, S.T., Slater, P.J., 1998a. Domination in Graphs: Advanced Topics. Marcel Dekker, New York.
Haynes, T.W., Hedetniemi, S.T., Slater, P.J., 1998b. Fundamentals of Domination in Graphs. Marcel Dekker, New York.
Hochbaum, D.S., 1982. SIAM J. Comput. 11, 555–556.
Janson, S., Łuczak, T., Rucinński, A., 2000. Random Graphs. Wiley, New York.
Jaromczyk, J.W., Toussaint, G.T., 1992. Relative neighborhood graphs and their relatives. Proceedings of the IEEE 80, 1502–1517.
Johnson, D.S., 1974. Approximate algorithms for combinatorical problems. J. Comput. System Sci. 9, 256–278.
Karoński, M., Scheinerman, E.R., Singer-Cohen, K.B., 1999. On random intersection graphs: the subgraph problem. Comb., Prob. Comput. 8, 131–159.
Karp, R., 1972. Reducibility among combinatorial problems. Plenum press, New York.
McKee, T.A., McMorris, F.R., 1999. Topics in Intersection Graph Theory. SIAM, Philadelphia, PA.
McMorris, F.R., Wang, C., 2000. Sphere-of-attraction graphs. Congr. Numer. 142, 149–160.
Nikoletseas, S.E., Spirakis, P.G., 1994. Near-optimal dominating sets in dense random graphs in polynomial expected time. Lecture Notes in Computer Science 790.
Parekh, A.K., 1991. Analysis of a greedy heuristic for finding small dominating sets in graphs. Inform. Process. Lett. 39, 237–240.
Priebe, C.E., Marchette, D.J., 2000. Characterizing the complexity of a high-dimensional classification problem. Comput. Sci. Statist. 32, 48–62.
Prisner, E., 1989. A characterization of interval catch digraphs. Discrete Math. 73, 285–289.
Prisner, E., 1994. Algorithms for interval catch digraphs. Discrete Appl. Math. 51, 147–157.
Sen, M., Das, S., Roy, A.B., West, D.B., 1989. Interval digraphs: An analogue of interval graphs. J. Graph Theory 13, 189–202.
Tuza, Z., 1994. Inequalities for minimal covering sets in set systems of given rank. Discrete Appl. Math. 51, 187–195.