15

Epsilon Nets and Transversals of Hypergraphs

For historical reasons, a finite set-system is often called a hypergraph. More precisely, a *hypergraph* H consists of a finite set $V(H)$ of *vertices* (points) and a family $E(H)$ of subsets of $V(H)$. The elements of $E(H)$ are usually called *hyperedges* (or, in short, *edges*). If the hyperedges of H are r-element sets, then H is said to be an *r-uniform hypergraph*. Using this terminology, a graph is a two-uniform hypergraph. In Chapter 10 we have extended some graph-theoretic results to r-uniform hypergraphs (cf. Theorems 10.11 and 10.12).

The concept of hypergraphs is a very general one, so it is not surprising that hypergraph theory has a large scale of applications in various fields of mathematics, including geometry. Given a hypergraph H, a subset $T \subseteq V(H)$ is called a *transversal* of H if $T \cap E$ is nonempty for every edge $E \in E(H)$. Many extremal problems from combinatorics and geometry can be reformulated as questions of the following type: What is the size of a smallest transversal in a given hypergraph H? This problem, in general, is known to be computationally intractable (cf. Garey and Johnson, 1979). However, under certain specific conditions on H, one can guarantee the existence of a relatively small transversal. The present chapter focuses on results of this kind. In particular, we shall see how a powerful probabilistic idea of Vapnik and Chervonenkis can be applied to obtain a number of interesting geometric and algorithmic results.

TRANSVERSALS AND FRACTIONAL TRANSVERSALS

Let H be a hypergraph with vertex set $V(H)$ and edge set $E(H)$. Let $\tau(H)$ denote the size of a smallest transversal of H, that is, the smallest number τ such that one can choose τ vertices with the property that any edge of H contains at least one of them. $\tau(H)$ is usually called the *transversal number* (or the *vertex-cover number*) of H.

The *packing number* (or *matching number*) of a hypergraph H is defined as the largest number $\nu = \nu(H)$ such that H has ν pairwise disjoint hyperedges. Obviously, $\nu(H) \le \tau(H)$ for any hypergraph H. Typically, $\tau(H)$ is strictly larger than $\nu(H)$. In fact, $\tau(H)$ cannot even be bounded by any function of $\nu(H)$ (see Exercise 15.3).

Let $\mathbb{R}^+$ denote the set of all nonnegative real numbers. Let us call a function $t: V(H) \rightarrow \mathbb{R}^+$ a *fractional transversal* of H if

$$\sum_{x \in E} t(x) \geq 1 \qquad \text{for every hyperedge } E \in E(H). \tag{15.1}$$

The minimum of $\sum_{x \in V(H)} t(x)$ over all fractional transversals of H is called the *fractional transversal number* of H, and is denoted by $\tau^*(H)$. One can associate with each transversal T of H a function $t_T: V(H) \rightarrow \mathbb{R}^+$ defined as

$$t_T(x) = \begin{cases} 1 & \text{if } x \in T, \\ 0 & \text{if } x \notin T. \end{cases}$$

Since this function satisfies (15.1) and $\sum_{x \in V(G)} t_T(x) = |T|$, we have that $\tau^*(H) \leq \tau(H)$.

Similarly, a *fractional packing* of H is a nonnegative function $p: E(H) \rightarrow \mathbb{R}^+$ such that

$$\sum_{x \in E} p(E) \leq 1 \qquad \text{for every vertex } x \in V(H).$$

The maximum of $\sum_{E \in E(H)} p(E)$ over all fractional packings of H is called the *fractional packing number* of H, and is denoted by $\nu^*(H)$. As before, we have $\nu^*(H) \geq \nu(H)$.

It is easy to deduce directly from the definition that $\nu^*(H) \leq \tau^*(H)$ (see Exercise 15.1). In fact, these two numbers are always equal to each other. Moreover, the following is true.

Theorem 15.1. *For every hypergraph H,*

$$\nu(H) \leq \nu^*(H) = \tau^*(H) \leq \tau(H),$$

and the value of $\nu^*(H) = \tau^*(H)$ can be determined by linear programming.

Proof. Let x_i $(1 \leq i \leq n)$ and E_j $(1 \leq j \leq m)$ be the vertices and the edges of H, respectively. Let $A = (a_{ij})$ be the *incidence matrix* of H, i.e.,

$$a_{ij} = \begin{cases} 1 & \text{if } x_i \in E_j, \\ 0 & \text{if } x_i \notin E_j. \end{cases}$$

Let A^T denote the transpose of A, and let $\mathbf{1}_n$ denote the matrix consisting of one column of length n, all of whose entries are 1's. Given a function $t: V(H) \rightarrow \mathbb{R}$ (and $p: E(H) \rightarrow \mathbb{R}$), let $\underline{t}$ (resp. $\underline{p}$) denote a matrix consisting of one column whose ith entry is $t(x_i)$ (resp. $p(E_i)$).

Observe that t is a fractional transversal of H if and only if

$$A^T\underline{t} \geq \mathbf{1}_m \quad \text{and} \quad \underline{t} \geq \mathbf{0}.$$

Similarly, p is a fractional packing of H if and only if

$$A\underline{p} \leq \mathbf{1}_n \quad \text{and} \quad \underline{p} \geq \mathbf{0}.$$

Thus,

$$\tau^*(H) = \min\{\mathbf{1}_n^T\underline{t} \,|\, A^T\underline{t} \geq \mathbf{1}_m, \underline{t} \geq \mathbf{0}\},$$
$$\nu^*(H) = \max\{\mathbf{1}_m^T\underline{p} \,|\, A\underline{p} \leq \mathbf{1}_n, \underline{p} \geq \mathbf{0}\}.$$

These two linear programming problems are dual to each other, so it follows immediately from the duality theorem of linear programming that their solutions, $\tau^*(H)$ and $\nu^*(H)$, are equal (see, e.g., Papadimitriou and Steiglitz, 1982; Chvátal, 1983; and Grötschel et al., 1987). □

In general, $\tau^*(H)$ can be much smaller than $\tau(H)$ (see Exercise 15.3). The following theorem of Lovász (1975) shows that this is not the case when every point of H belongs to relatively few hyperedges.

Theorem 15.2 (Lovász). *Let H be a hypergraph whose every vertex is contained in at most D edges. Then*

$$\tau^*(H) \leq \tau(H) \leq (\ln D + 1)\tau^*(H).$$

Proof. We have to prove only the second inequality. Let $t: V(H) \to \mathbb{R}^+$ be a fractional transversal of H with $\sum_{x \in V(H)} t(x) = \tau^*(H)$.

We are going to select a set of vertices $x_1, x_2, \ldots$ by a *greedy* algorithm. Let x_1 be any vertex of H whose degree (i.e., the number of edges containing it) is maximal. Let D_1 denote the degree of x_1 in H. Set $H_1 = H - x_1$, that is, the hypergraph obtained from H by deleting the vertex x_1 and all edges containing x_1. If $x_1, \ldots, x_i \in V(H)$ have already been selected, then let $H_i = H - x_1 - x_2 - \cdots - x_i$. If H_i has no edges, we stop. Otherwise, let x_{i+1} be a vertex of H_i whose degree D_{i+1} is maximal, and so on. Clearly,

$$|E(H_i)| - |E(H_{i+1})| = D_{i+1}. \tag{15.2}$$

By the properties of t,

$$\begin{aligned}
|E(H_i)| &= \sum_{E \in E(H_i)} 1 \le \sum_{E \in E(H_i)} \sum_{x \in E} t(x) \\
&= \sum_{x \in V(H_i)} t(x) \sum_{\substack{E \in E(H_i) \\ E \ni x}} 1 \\
&\le \sum_{x \in V(H_i)} t(x) D_{i+1} \\
&\le D_{i+1} \tau^*(H).
\end{aligned}$$

Assume now that our procedure terminates in s steps; i.e., H_s is empty. Then, of course, $\tau(H) \le s$. Put $H_0 = H$. By (15.2), we have

$$\begin{aligned}
s = \sum_{i=0}^{s-1} 1 &= \sum_{i=0}^{s-1} \frac{|E(H_i)| - |E(H_{i+1})|}{D_{i+1}} \\
&= \frac{|E(H)|}{D_1} + \sum_{i=1}^{s-1} |E(H_i)| \Big(\frac{1}{D_{i+1}} - \frac{1}{D_i} \Big).
\end{aligned}$$

Hence, using the inequality $|E(H_i)| \le D_{i+1}\tau^*(H)$ $(0 \le i \le s)$ and the fact $D_s \ge 1$, we obtain

$$\begin{aligned}
s &\le \tau^*(H) + \sum_{i=1}^{s-1} D_{i+1}\tau^*(H) \Big(\frac{1}{D_{i+1}} - \frac{1}{D_i} \Big) \\
&= \tau^*(H) \Big(1 + \sum_{i=1}^{s-1} \frac{D_i - D_{i+1}}{D_i} \Big) \\
&\le \tau^*(H) \Big(1 + \sum_{i=1}^{s-1} \sum_{k=D_{i+1}+1}^{D_i} \frac{1}{k} \Big) \\
&= \tau^*(H) \Big(1 + \sum_{k=D_s+1}^{D_1} \frac{1}{k} \Big) \\
&\le \tau^*(H) (1 + \ln D_1).
\end{aligned}$$

Thus,

$$\tau(H) \le s \le \tau^*(H) (1 + \ln D_1),$$

as desired. □

VAPNIK–CHERVONENKIS DIMENSION

Suppose that for a public opinion poll we want to select a small number of individuals representing all major sections of the society. First, we have to choose certain categories of people and then decide which of these groups are considered "important." According to our democratic principles, we shall measure the "importance" of a group by its size (in the percentage of the population). Then the important groups will define a hypergraph H with the property that $|E| \geq \varepsilon|V(H)|$ for every edge $E \in E(H)$, where ε is some fixed constant $(0 < \varepsilon < 1)$. The smallest number of people representing all important groups is $\tau(H)$.

Clearly, the function $t(x) = 1/(\varepsilon|V(H)|)$, for all $x \in V(H)$, is a fractional transversal of H with $\sum_{x \in V(H)} t(x) = 1/\varepsilon$. Hence, $\tau^*(H) \leq 1/\varepsilon$, and Theorem 15.2 implies that

$$\tau(H) \leq \frac{1}{\varepsilon}(\ln D + 1), \tag{15.3}$$

where D is the maximum degree of the vertices of H. This bound is extremely poor if D is large.

In their seminal paper, Vapnik and Chervonenkis (1971) pointed out that if H satisfies certain natural conditions, the above upper bound can be replaced by a function depending only on ε. To specify these conditions, we need some preparation.

Definition 15.3. Let $H = (V(H), E(H))$ denote a hypergraph. A subset $A \subseteq V(H)$ is called *shattered* if for every $B \subseteq A$ there exists an $E \in E(H)$ such that $E \cap A = B$. The *Vapnik–Chervonenkis dimension* (or *VC dimension*) of H is the cardinality of the largest shattered subset of $V(H)$. It will be denoted by VC-dim(H).

The following theorem was proved independently by Shelah (1972), Sauer (1972), and Vapnik and Chervonenkis (1971).

Theorem 15.4. *Let H be a hypergraph with n vertices and VC-dimension d. Then*

$$|E(H)| \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d},$$

and this bound cannot be improved.

First Proof. The assertion is trivial if $d = 0$ or $n \leq d$. Assume that we have already proved it for every hypergraph $\overline{H}$ with VC-dim$(\overline{H}) < d$, and for every hypergraph $\overline{H}$ with VC-dim$(\overline{H}) = d$ and $|V(\overline{H})| < n$.

Given a hypergraph H with n vertices and VC-dimension d, let us define two other hypergraphs, H_1 and H_2, as follows. Let $V(H_1) = V(H_2) = V(H) - \{x\}$ for some fixed $x \in V(H)$, and set

$$E(H_1) = \{E - \{x\} \mid E \in E(H)\},$$
$$E(H_2) = \{E \in E(H) \mid x \notin E \quad \text{and} \quad E \cup \{x\} \in E(H)\}.$$

Obviously, VC-dim $(H_1) \le d$ and VC-dim $(H_2) \le d - 1$.

On the other hand, by the induction hypothesis,

$$\begin{aligned} |E(H)| &= |E(H_1)| + |E(H_2)| \\ &\le \sum_{i=0}^{d} \binom{n-1}{i} + \sum_{i=0}^{d-1} \binom{n-1}{i} \\ &= \sum_{i=0}^{d} \binom{n}{i}. \end{aligned}$$

The tightness of this bound follows from the fact that if $E(H) = \{U \subseteq V \mid |U| \le d\}$, then VC-dim $(H) = d$. □

We also include a slightly more complicated proof due to Frankl and Pach (1983), because it is a good illustration of the so-called *linear algebra method* (see, e.g., Babai and Frankl, 1988).

Second Proof. Let $E(H) = \{E_i \mid 1 \le i \le m\}$, and let X_j, $1 \le j \le \sum_{i=0}^{d} \binom{n}{i}$, be a list of all subsets of $V(H)$ of size at most d. Define an $m \times \sum_{i=0}^{d} \binom{n}{i}$ matrix $\mathbf{A} = (a_{ij})$ by

$$a_{ij} = \begin{cases} 1 & \text{if } E_i \supseteq X_j, \\ 0 & \text{if } E_i \not\supseteq X_j. \end{cases}$$

Suppose, for contradiction, that $m > \sum_{i=0}^{d} \binom{n}{i}$. Then the rows of $\mathbf{A}$ are linearly dependent over the reals; thus there exists a nonzero function $f: E(H) \to \mathbb{R}$ such that

$$\sum_{E_i \supseteq X_j} f(E_i) = 0 \qquad \text{for every } X_j.$$

Let $A \subseteq V(H)$ be a *minimal* subset for which

$$\sum_{E_i \supseteq A} f(E_i) = \alpha \ne 0.$$

(Sets A with nonzero sums certainly exist, for we get a nonzero sum for any maximal element A of the family $\{A \in E(H) | f(A) \ne 0\}$.) Obviously, $|A| \ge d + 1$. Given any $B \subseteq A$, let

$$F(B) = \sum_{E_i \cap A = B} f(E_i).$$

Thus, $F(A) = \alpha$, and setting $B = A - \{a\}$ for any fixed $a \in A$,

$$\begin{aligned} F(B) &= \sum_{E_i \supseteq B} f(E_i) - \sum_{E_i \supseteq A} f(E_i) \\ &= 0 - \alpha = -\alpha \end{aligned}$$

In general, if B is any $(|A| - k)$-element subset of A $(0 \le k \le |A|)$, then

$$F(B) = (-1)^k \alpha \ne 0.$$

This yields, in particular, that there exists at least one hyperedge E_i with $E_i \cap A = B$. Thus, A is shattered, contradicting our assumption that VC-dim$(H) = d$. □

Vapnik and Chervonenkis (1971) discovered an ingenious probabilistic (counting) argument based on the above result, which leads to a substantial improvement of the bound (15.3). They showed (in a somewhat different setting) that there exists a function $f(d, \varepsilon)$ such that the transversal number of every hypergraph H of VC-dimension d, all of whose edges have at least $\varepsilon|V(H)|$ elements, is at most $f(d, \varepsilon)$ (see Exercise 15.6). The ideas of Vapnik and Chervonenkis have been adapted by Haussler and Welzl (1987) and Blumer et al. (1989) to obtain various upper bounds on $f(d, \varepsilon)$. These results were sharpened and generalized by Komlós, Pach, and Woeginger (1992), as follows. Given a finite set V, a function $\mu: V \to \mathbb{R}^+$ is called a *probability measure* if

$$\sum_{x \in V} \mu(x) = 1.$$

The measure of any subset $X \subseteq V$ is defined by $\mu(X) = \sum_{x \in X} \mu(x)$.

Theorem 15.5 (Komlós et al.). *Let H be a hypergraph of VC-dimension d, let $\varepsilon > 0$, and let μ be a probability measure on $V(H)$ such that $\mu(E) \ge \varepsilon$ for every $E \in E(H)$. Then $\tau(H) \le t(d, \varepsilon)$, where $t(d, \varepsilon)$ denotes the smallest positive integer t satisfying*

$$2\sum_{i=0}^{d} \binom{T}{i} \left(1 - \frac{t}{T}\right)^{(T-t)\varepsilon - 1} < 1$$

for some integer $T > t$. Consequently, for any $\varepsilon \le \frac{1}{2}$, we have

$$\tau(H) \le \frac{d}{\varepsilon}\left(\ln \frac{1}{\varepsilon} + 2 \ln \ln \frac{1}{\varepsilon} + 6\right)$$

(cf. Exercise 15.9).

Proof. Let us select with possible repetition t random points of $V(H)$, where the selections are done with respect to the probability measure μ. We get a *random sample*

$$x \in [V(H)]^t = \underbrace{V(H) \times \cdots \times V(H)}_{t\text{times}}.$$

We say that x is a *transversal* of H if every edge $E \in E(H)$ contains at least one point of x. Let $I(E, x)$ denote the number of components of x that belong to E, counting with multiplicity. Then

$$\Pr[x \text{ is not a transversal of } H] = \Pr[\exists E \in E(H) : I(E, x) = 0].$$

Having picked the string x of length t, let us choose randomly another $T - t$ elements from $V(H)$. Let $y \in [V(H)]^{T-t}$ denote this new string, and let $z = xy \in [V(H)]^T$ stand for the full sequence. Furthermore, let $\langle z \rangle = \langle xy \rangle$ denote the *multiset* of all elements occurring in z (i.e., they are counted with multiplicities but their order is irrelevant).

For any $E \in E(H)$, $I(E, y)$ is a random variable having binomial distribution. Let m_E be the *median* of $I(E, y)$,

$$\Pr[I(E, y) > m_E] \le \tfrac{1}{2} \le \Pr[I(E, y) \ge m_E].$$

The following inequality is an immediate consequence of the independence of x and y.

$$
\begin{aligned}
&\Pr[\exists E \in E(H) : I(E, x) = 0] \\
&\le \frac{\Pr[\exists E \in E(H) : I(E, x) = 0 \text{ and } I(E, y) \ge m_E]}{\min\limits_{E \in E(H)} \Pr[I(E, y) \ge m_E]}. \\
&\le 2 \Pr[\exists E \in E(H) : I(E, x) \\
&= 0 \text{ and } I(E, y) \ge m_E].
\end{aligned}
$$

For a fixed $E \in E(H)$, the conditional probability for given $\langle z \rangle = \langle xy \rangle$

$$\begin{aligned}
\Pr\,[I(E,x)=0 \text{ and } I(E,y)\ge m_E \,|\, \langle z\rangle]
&= \chi\,[I(E,z)\ge m_E]\,\frac{\binom{T-t}{I(E,z)}}{\binom{T}{I(E,z)}}\\
&\le \chi\,[I(E,z)\ge m_E]\left(1-\frac{t}{T}\right)^{I(E,z)}\\
&\le \chi\,[I(E,z)\ge m_E]\left(1-\frac{t}{T}\right)^{m_E}.
\end{aligned}$$

(Here $\chi\,[A]$ is the characteristic function of A, that is, $\chi\,[A]=1$ if A is true, and 0 otherwise.)

By Theorem 15.4, a fixed multiset $\langle z\rangle$ has at most $\sum_{i=0}^{d}\binom{T}{i}$ different intersections with the edges of H. Thus,

$$\Pr\,[\exists\, E\in E(H): I(E,x)=0 \text{ and } I(E,y)\ge m_E \,|\, \langle z\rangle] \le \left(\sum_{i=0}^{d}\binom{T}{i}\right)\left(1-\frac{t}{T}\right)^{m},$$

where $m=\min_{E\in E(H)} m_E$. Using the known fact that the median of a binomial distribution is within 1 of the mean,

$$m \ge (T-t)\min_{E\in E(H)} \mu(E)-1 \ge (T-t)\varepsilon-1.$$

Hence, we obtain

$$\Pr\,[\exists\, E\in E(H): I(E,x)=0] \le 2\left(\sum_{i=0}^{d}\binom{T}{i}\right)\left(1-\frac{t}{T}\right)^{(T-t)\varepsilon-1}.$$

If the last expression is less than 1, then x is a transversal of H, with positive probability. This proves the first statement of the theorem. Choosing

$$t=\left\lfloor \frac{d}{\varepsilon}\left(\ln\frac{1}{\varepsilon}+2\ln\ln\frac{1}{\varepsilon}+6\right)\right\rfloor,$$

$$T=\left\lfloor \frac{\varepsilon}{d}t^2\right\rfloor,$$

we get after some calculations that

$$2\sum_{i=0}^{d}\binom{T}{i}\left(1-\frac{t}{T}\right)^{(T-t)\varepsilon-1}<1,$$

provided that $\varepsilon\le\frac{1}{2}$. □

The above theorem is valid for any probability measure μ defined on the vertex set of H. In particular, one can choose μ to be constant; that is, $\mu(x) = 1/|V(H)|$ for every $x \in V(H)$. We can deduce another interesting result from Theorem 15.5 by applying it to the measure $\mu'(x) = t(x)/\tau^*(H)$, where $t: V(H) \to \mathbb{R}^+$ is a fractional transversal of H with $\sum_{x \in V(H)} t(x) = \tau^*(H)$. Observe that in this case

$$\begin{aligned}\mu'(E) &= \sum_{x \in E} \mu'(x)\\ &= \sum_{x \in E} \frac{t(x)}{\tau^*(H)} \geq \frac{1}{\tau^*(H)}\end{aligned}$$

holds for every $E \in E(H)$. Thus, choosing $\varepsilon = 1/\tau^*(H)$ in Theorem 15.5, we obtain the following.

Corollary 15.6 (Komlós et al.). *Let H be any hypergraph of VC-dimension d.*

(i) *If every edge of H has at least $\varepsilon|V(H)|$ elements for some $\varepsilon \leq \frac{1}{2}$, then*

$$\tau(H) \leq \frac{d}{\varepsilon}\left(\ln \frac{1}{\varepsilon} + 2 \ln \ln \frac{1}{\varepsilon} + 6\right).$$

(ii) *If $\tau^*(H) \geq 2$, then*

$$\tau(H) \leq d\tau^*(H)(\ln \tau^*(H) + 2 \ln \ln \tau^*(H) + 6).$$

Next we show that for $d \geq 2$, the bound given in Theorem 15.5 is close to being optimal.

Theorem 15.7 (Komlós et al.). *Given any natural number $d \geq 2$ and any real $\gamma < 2/(d+2)$, there exists a constant $\varepsilon_{d,\gamma} > 0$ with the following property.*

For any $\varepsilon \leq \varepsilon_{d,\gamma}$, one can construct a hypergraph H of VC-dimension d, all of whose edges have at least $\varepsilon|V(H)|$ points, and

$$\tau(H) \geq (d - 2 + \gamma)\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}.$$

Proof. Again, we use the probabilistic method. Let γ' be a fixed constant, $\gamma < \gamma' < 2/(d+2)$. Given a sufficiently small ε, let $n = (K/\varepsilon)\ln(1/\varepsilon)$, where K is a constant depending only on d, γ, and γ' (but not on ε), which will be specified later. Furthermore, let

$$r = \varepsilon n, \qquad p = \frac{\varepsilon^{1-d-\gamma'}}{\binom{n}{r}}, \qquad t = (d - 2 + \gamma)\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}.$$

We assume that n, r, t are integers, disregarding all roundoff errors.

Let V be a fixed n-element set. Construct a hypergraph H on the vertex set V by randomly selecting some r-element subsets of V, where each r-tuple is chosen independently with probability p. We are going to show that with high probability

(i) $\text{VC-dim}\,(H) \le d$, and

(ii) $\tau(H) > t$.

$$
\begin{aligned}
&\Pr[\text{VC-dim}\,(H) > d] \\
&\quad \le \binom{n}{d+1} \Pr[\text{a fixed } (d+1)\text{-element subset } A \subseteq V \text{ is shattered by } H] \\
&\quad = \binom{n}{d+1} \prod_{B \subseteq A} \Pr[\exists E \in E(H) : E \cap A = B] \\
&\quad = \binom{n}{d+1} \prod_{B \subseteq A} \left(1 - (1-p)^{\binom{n-|A|}{r-|B|}}\right) \\
&\quad = \binom{n}{d+1} \prod_{j=0}^{d+1} \left(1 - (1-p)^{\binom{n-d-1}{r-j}}\right)^{\binom{d+1}{j}} \\
&\quad = \binom{n}{d+1} \prod_{i=0}^{d+1} \left(1 - (1-p)^{\binom{n-d-1}{r-d-1+i}}\right)^{\binom{d+1}{d+1-i}} \\
&\quad = \binom{n}{d+1} \prod_{i=0}^{d+1} \left(1 - (1-p)^{\binom{n-d-1}{r-d-1+i}}\right)^{\binom{d+1}{i}} \\
&\quad \le \binom{n}{d+1} \prod_{i=0}^{1} \left(p \binom{n-d-1}{r-d-1+i}\right)^{\binom{d+1}{i}} \\
&\quad = \binom{n}{d+1} p \binom{n-d-1}{r-d-1} \left(p \binom{n-d-1}{r-d}\right)^{d+1} \\
&\quad \le n^{d+1} p \binom{n}{r} \left(\frac{r}{n}\right)^{d+1} \left(p \binom{n}{r} \left(\frac{r}{n}\right)^{d}\right)^{d+1} \\
&\quad = \left(K \ln \frac{1}{\varepsilon}\right)^{d+1} \varepsilon^{2-(d+2)\gamma'},
\end{aligned}
$$

which tends to 0 as $\varepsilon \to 0$. This proves (i).

Next we show that (ii) also holds with high probability.

$$\begin{aligned}\Pr[\tau(H) \le t] &= \binom{n}{t}(1-p)^{\binom{n-t}{r}} \\ &\le \binom{n}{t}\exp\left[-p\binom{n-t}{r}\right] \\ &\le \left(\frac{en}{t}\right)^t \exp\left[-p\binom{n}{r}\left(1-\frac{r}{n-t+1}\right)^t\right].\end{aligned}$$

From this, using the inequality $1 - ax > e^{-bx}$ for $b > a$, $0 < x < 1/a - 1/b$, we obtain the upper bound

$$\begin{aligned}&\left(\frac{en}{t}\right)^t \exp\left[-p\binom{n}{r}e^{-K\varepsilon t/(K-d)}\right] \\ &= \left(\frac{eK}{d-2+\gamma}\right)^t \exp\left[-\varepsilon^{1-d-\gamma'+K(d-2+\gamma)/(K-d)}\right],\end{aligned}$$

which tends to 0 if

$$1 - d - \gamma' + K(d-2+\gamma)/(K-d) < -1,$$

i.e., if K is sufficiently large. □

The condition $d \ge 2$ in Theorem 15.7 is not merely a technical assumption. In fact, it is not hard to characterize all finite hypergraphs H with VC-dimension 1, and one can check that $\tau(H) \le \lceil 1/\varepsilon \rceil - 1$, provided that every edge of H has at least $\varepsilon|V(H)|$ points for some $0 < \varepsilon < 1$ (see Exercise 15.8).

The following simple assertion will help us in deciding whether a given hypergraph has low VC-dimension.

Lemma 15.8. *Let H be a hypergraph of VC-dimension d, and let $\varphi(E_1, \dots, E_k)$ be a set-theoretic formula of k variables (using $\cup, \cap, -$). If every edge E' of a hypergraph H' can be expressed as*

$$E' = \varphi(E_1, \dots, E_k) \qquad \text{for suitable } E_i \in E(H),$$

then

$$\text{VC-dim}(H') \le 2dk\log(2dk).$$

Proof. Let A be a d'-element subset of $V(H') = V(H)$, which is shattered in H'. By Theorem 15.4,

$$|\{E \cap A \mid E \in E(H)\}| \le \sum_{i=0}^{d}\binom{d'}{i}.$$

Using the assumption on H', this yields

$$2^{d'} = |\{E' \cap A \mid E' \in E(H')\}| \le \Big(\sum_{i=0}^{d} \binom{d'}{i} \Big)^k.$$

Comparing the two sides of this inequality, we obtain that $d' \le 2dk \log(2dk)$, as required. □

Ding, Seymour, and Winkler (1994) have introduced another parameter of a hypergraph, closely related to its VC-dimension. They defined $\lambda(H)$ as the largest integer l such that one can choose l edges $E_1, E_2, \ldots, E_l \in E(H)$ with the property that for any $1 \le i < j \le l$, there is a vertex $x_{ij} \in E_i \cap E_j$ that does not belong to any other E_g $(g \ne i, j)$. It is easy to see that VC-dim$(H) < \binom{\lambda(H)+1}{2}$ for every hypergraph H. Combining Corollary 15.6 with Ramsey's theorem (Theorem 9.13), one can establish the following result.

Theorem 15.9 (Ding et al., 1994). *For any hypergraph H,*

$$\tau(H) \le 6\lambda^2(H)\,(\lambda(H) + \nu(H)) \binom{\lambda(H) + \nu(H)}{\lambda(H)}^2.$$

At the beginning of this chapter we pointed out that in general it is impossible to bound τ from above by any function of ν. Gyárfás and Lehel (1983, 1985) initiated the investigation of certain classes of hypergraphs for which such functions exist. Theorem 15.9 provides a sufficient condition for a family of hypergraphs to have this property. It implies that if there exists a constant K such that $\lambda(H) \le K$ for all members of a family, then τ can be bounded from above by a polynomial of ν. For various geometric consequences of this fact, see Pach (1995).

RANGE SPACES AND ε-NETS

Haussler and Welzl (1987) were the first to recognize the relevance of the above machinery to geometric problems, and in fact they formulated and proved the first version of Theorem 15.5, too. It seems to capture the essence of the so-called *random* (or *probabilistic*) method in a large variety of geometric applications. This ready-to-use kit will save us a lot of time (and space) in situations where otherwise we would go through lengthy but routine calculations. However, the main significance of these ideas is that they shed some light on the general transversal problem. The transversal number is a *global* parameter of a set system. The results in the preceding section show that in any measure space of total measure 1, any system of large measurable sets admits a relatively small transversal, provided that its *local* behavior is nice (i.e., its VC-dimension is bounded).