# The Vapnik-Chervonenkis Dimension of a Random Graph

Martin Anthony* Graham Brightwell† Colin Cooper‡

**Abstract**

In this paper we investigate a parameter defined for any graph, known as the *Vapnik-Chervonenkis dimension* (or VC dimension). For any vertex $x$ of a graph $G$, the closed neighbourhood $N(x)$ of $x$ is the set of all vertices of $G$ adjacent to $x$, together with $x$. We say that a set $D$ of vertices of $G$ is *shattered* if every subset $R$ of $D$ can be realised as $R = D \cap N(x)$ for some vertex $x$ of $G$. The Vapnik-Chervonenkis dimension of $G$ is defined to be the largest cardinality of a shattered set of vertices. Our main result gives, for each positive integer $d$, the exact threshold function for a random graph $G(n,p)$ to have VC dimension $d$.

*Department of Mathematics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE

†Department of Mathematics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE

‡School of Mathematical Studies, University of North London, 2-16 Eden Grove, London N7 8EA

# 1 Introduction

In this paper we investigate a parameter defined for any graph: the *Vapnik-Chervonenkis dimension*. The VC dimension of a graph was defined by Haussler and Welzl [7] and is an interesting special case of the more general and well-established notion of the Vapnik-Chervonenkis dimension of a set system, first introduced in [11]. The Vapnik-Chervonenkis dimension has proved useful in a number of areas of mathematics and computer science; in probability theory [11, 10, 8], in computational geometry [7] and in the theory of machine learning [4, 2], for example.

We start by presenting the necessary definitions and making a few preliminary observations. Our main aim is to determine, for each positive integer $d$, the exact edge-probabililty threshold function for a random graph $G(n,p)$ to have VC dimension at least $d$: for large $d$, this turns out to be about $p = n^{-1/d}$. The authors are currently working on another paper dealing with the VC dimension of a random graph $G(n,p)$ for larger values of $p = p(n)$. The problem of estimating the VC dimension of a random graph was first suggested by Colin McDiarmid at the 1991 British Combinatorial Conference.

# 2 Definitions and preliminaries

We start with some standard definitions. Suppose that $\mathcal{F}$ is a family of subsets of a finite set $X$. For $D \subseteq X$, the *trace* of $\mathcal{F}$ on $D$ is $\Pi_{\mathcal{F}}(D) = \{D \cap F : F \in \mathcal{F}\}$ and the subsets of $D$ of the form $D \cap F$ with $F \in \mathcal{F}$ are known as *dichotomies* (of $D$ by $\mathcal{F}$). A subset $D$ of $X$ is said to be *shattered* by $\mathcal{F}$ if $\Pi_{\mathcal{F}}(D) = 2^D$, the power set of $D$. In this case, each subset of $D$ can be realised as a dichotomy of $D$ by $\mathcal{F}$. The *Vapnik-Chervonenkis dimension* [11, 7] (or *VC dimension*) of $\mathcal{F}$, denoted $\mathrm{VCdim}(\mathcal{F})$, is the maximal $d$ such that *some* subset of $X$ of cardinality $d$ is shattered by $\mathcal{F}$.

Following Haussler and Welzl [7], we define the VC dimension of a graph as follows. Let $G = (V,E)$ be a (simple, loopless) graph with vertex-set $V = V(G)$ and edge-set $E = E(G)$. The *(closed) neighbourhood* of a vertex $v$ is the set

$$N(v) = \{u \in V : \{u,v\} \in E\} \cup \{v\},$$

the set of all vertices at distance at most 1 from $v$. Denote by $N(G)$ the set of all neighbourhoods of vertices of $G$,

$$N(G) = \{N(v) : v \in V\}.$$

Then $N(G)$, as a family of subsets of the set $V$, has a VC dimension, which we shall call the VC dimension of the graph $G$. Thus, a set $D$ of vertices is shattered (by $N(G)$) if every subset $R$ of $D$ can be realised as $R = D \cap N(x)$ for some vertex $x$ of $G$. In this case, we say that $R$ is *generated* by $x$, and that $x$ is a *generator* of $R$; if $x \in D$, we say that $x$ *internally generates* $R$, while if $x \notin D$, $x$ *externally generates* $R$.

A related parameter is the *testing dimension* of $G$, denoted $\mathrm{Tdim}(G)$, which is the maximal $d$ such that *all* subsets of $X$ of cardinality $d$ are shattered by $N(G)$. The testing dimension of a set system has been studied in several recent papers [9, 3]. We do not deal with the testing dimension of a graph in this paper, but it seems to us that it is also a natural object of study.

A graph $G = (V, E)$ is said to be *homeomorphic* to a graph $H$ if (an isomorphic copy of) $G$ can be obtained from $H$ by the addition and removal of vertices of degree two (the incidence being changed in the obvious manner). Haussler and Welzl [7] noted that if a graph has VC dimension at least 5 then it contains a subgraph homeomorphic to the complete graph on five vertices. (In particular, therefore, all planar graphs have VC dimension at most 4. In fact, it is easy to construct planar graphs with VC dimension 4.) More generally, we have the following straightforward result.

**Theorem 1** *If a graph $G$ has VC dimension at least $n$, then $G$ has a subgraph homeomorphic to the complete graph $K_n$ on $n$ vertices.*

**Proof:** Suppose that $S$ is a set of $n$ vertices of $G$ shattered by $N(G)$ and that $x, y \in S$ are not adjacent in $G$. Since $S$ is shattered, there is a vertex $w = w(x, y)$ such that $\{x, y\} = N(w) \cap S$. Since $x, y$ are non-adjacent, $w \neq x$ and $w \neq y$. Therefore $w$ is a vertex in $V \setminus S$ such that the only vertices of $S$ adjacent to $w$ are $x$ and $y$. This analysis holds for each pair of non-adjacent vertices in $S$. The subgraph $H$ formed by the edges inside $S$ and all the edges $xw$, $yw$, where $x, y$ are non-adjacent vertices of $S$ and $w = w(x, y)$, is thus homeomorphic to $K_n$. □

Note that, for this result to hold, it is not necessary that we have every subset of $S$ equal to a dichotomy of $S$ by $N(G)$; rather, all we require is that every 2-subset of $S$ be a dichotomy of $S$ by $N(G)$.

One easy observation can be made concerning the VC dimension of a graph. Since there are at most $n$ closed neighbourhoods of an $n$-vertex graph $G$, the trace of $N(G)$ on any set $D$ of vertices consists of at most $n$ distinct sets. It follows that if $D$ is shattered then $2^{|D|} \le n$ and so $\mathrm{VCdim}(G) \le \lfloor \log_2 n \rfloor$. (Here, and throughout, $\log_2$ denotes logarithm to base 2 and log denotes natural logarithm.) This bound is tight. Let $n$ be any positive integer and $k = \lfloor \log_2 n \rfloor$. Take a set $K$ of $k$ independent vertices. For each non-singleton subset $R$ of $K$, introduce a vertex $x_R$ adjacent to precisely the vertices of $R$. This construction results in a graph with at most $n$ vertices and VC dimension $k = \lfloor \log_2 n \rfloor$.

# 3 Threshold functions for fixed VC dimension

The model we use for random graphs is the standard $G(n,p)$ model [5], which is defined as follows: let $\mathcal{G}(n)$ be the set of all labelled (loopless, simple) graphs on $n$ vertices and define the probability measure $\mu$ on the set of all subsets of $G(n,p)$ by specifying that a graph $H$ with $e$ edges has $\mu(H) = p^e(1-p)^{N-e}$, where $N = \binom{n}{2}$. We say that *almost every* $G(n,p)$ has a property $\Pi$, or that the random graph $G(n,p)$ *almost surely* has property $\Pi$, if, as $n \to \infty$, the $\mu$-probability of the set of graphs in $\mathcal{G}(n)$ having $\Pi$ tends to 1.

We are interested in determining when a random graph $G(n,p)$ almost surely has VC dimension at least $d$, where $d$ is a fixed integer.

It is intuitively clear that, as $p = p(n)$ increases from 0 to 1, the VC dimension of $G(n,p)$ starts off at 1, rises until $p \simeq 1/2$, then falls to 0 at $p = 1$. (Note the asymmetry, caused by the fact that we are dealing with *closed* neighbourhoods.) For $p < 1/2$, the main obstruction to having large VC dimension will be the scarcity of edges; whereas for $p > 1/2$, the main obstruction is the scarcity of non-edges. We deal only with the case of $p \le 1/2$, although it is not hard to adapt our arguments to deal with the other case. In fact, for this paper, we deal only with very small values of $p$: somewhat larger values of $p$, in particular constant $p$, will be dealt with in a subsequent paper.

Let $\Pi$ be a property of graphs. We say that $f(n)$ is a *threshold function* for $\Pi$ if, whenever $p/f(n) \to 0$ as $n \to \infty$, almost every $G(n,p)$ does not have $\Pi$, whereas, if, as $n \to \infty$, $p/f(n) \to \infty$ with $p \le 1/2$, then almost every $G(n,p)$ does have $\Pi$.

The *average degree* $ad(H)$ of a graph $H$ is $2e(H)/v(H)$, where $e(H)$ is the number of edges and $v(H)$ is the number of vertices of $H$. In what follows, for convenience, if $F$ is a subgraph of $H$, we write $F \subseteq H$. For a given graph $H$, the *maximum average degree* is the average degree of its densest subgraph. In other words, the maximum average degree, $m(H)$, of $H$ is the maximum of $ad(F)$, taken over all subgraphs $F$ of $H$. It is known [5], Chapter IV, that the maximum average degree of $H$ determines a threshold function $f(n)$ for $G(n,p)$ to contain $H$, either as a subgraph or as an induced subgraph.

**Theorem 2** *Let $H$ be a graph with maximum average degree $m$. Then the function $f(n) = n^{-2/m}$ is a threshold function for the property that a graph contains a subgraph isomorphic to $H$. The same function $f(n)$ is a threshold function for the property that a graph contains an induced subgraph isomorphic to $H$.*

It follows immediately from Theorem 2 that, if $\{H_1, \ldots, H_k\}$ is a finite set of graphs, with $\min m(H_i) = \mu$, then the function $f(n) = n^{-2/\mu}$ is a threshold function for the property that a graph contains one of the $H_i$ as an (induced) subgraph.

For a fixed natural number $d$, we define the class $\mathcal{H}_d$ of *d-shattering graphs* as follows. A graph $H$ in $\mathcal{H}_d$ induces some graph on a *base set* $D$ of $d$ vertices. Then, for each subset $S$ of $D$ not internally generated by a vertex of $D$, there is exactly one vertex $v_S$ of $H$ outside $D$ adjacent only to the vertices of $S$.

Note that, if $H$ is a $d$-shattering graph with base set $D$, then every edge of $H$ is incident with a vertex of $D$: we say that $H$ is *based on* $D$. Also, $H$ is determined by its restriction to $D$. Thus there are at most $2^{\binom{d}{2}}$ non-isomorphic $d$-shattering graphs, each with between $2^d$ and $2^d + d - 1$ vertices.

The relevance of $d$-shattering graphs is brought out by the following two immediate observations. First, if a graph $G$ has VC dimension at least $d$, then it contains some $d$-shattering graph as a subgraph. Conversely, if $G$ contains a $d$-shattering graph as an *induced* subgraph, then its VC dimension is at least $d$.

By Theorem 2 and the remark after, a threshold function for having VC dimension at least $d$ is thus $p(n) = n^{-2/\mu}$, where

$$\mu = \mu(d) = \min\{m(H) : H \in \mathcal{H}_d\} = \min_{H \in \mathcal{H}_d} \max_{F \subseteq H} ad(F).$$

In other words, to find this threshold function, we need to identify the *minimal* maximum average degree of a $d$-shattering graph.

We shall show that the problem of finding a $d$-shattering subgraph $H$ based on $D$ minimising $m(H)$ reduces to comparing two graphs $E$ and $I$ based on $D$, each of which is, in a sense, 'extreme'. The restriction of $E$ to $D$ is the empty graph, and the restriction of $I$ to $D$ is a graph $F(D)$ such that the sum of the cardinalities of the distinct subsets of $D$ generated by vertices of $D$ is maximised.

Let $G = (V, E)$ be a graph and $D$ be a subset of $V$. Let $X$ be a set of generators for some specified family $\mathcal{B}$ of subsets of $D$, where, for each $B \in \mathcal{B}$, a unique $x \in X$ has been selected such that $B = N(x) \cap D$. (Note that $X$ and $D$ need not be disjoint.) The edges $xy$ such that $x \in X$ and $y \in N(x) \cap D$ are called the *generating edges*. Thus $X$ and the generating edges define a *generating subgraph* $H = (X \cup D, \{xy : x \in X, y \in N(x) \cap D\})$. We say $H$ *minimally generates* the family $\mathcal{B}$: note that this implies that $H$ is based on $D$.

Fix a subset $D$ of $d$ vertices and let $H$ be a graph which minimally generates the family $\mathcal{D}_s = \{B \in 2^D, |B| \geq s\}$, with a set $X = X(H)$ of generators. Then we have

$$|X| = |\mathcal{D}_s| = \sum_{i \geq s} \binom{d}{i} \equiv \alpha(s),$$

and therefore

$$|V(H)| = \alpha(s) + |D \setminus X|.$$

We first discuss two specific graphs $E_s = E_s(d)$ and $I_s = I_s(d)$, which minimally generate $\mathcal{D}_s$. We aim to show that, for each $d$, one of $E = E_0$ and $I = I_0$ is the $d$-shattering graph with minimal maximum average degree, and that the subgraph of $E$ and $I$ with the largest maximum degree is of the form $E_s$ or $I_s$ respectively.

For $0 \leq s \leq d$, let $E_s$ be the graph based on $D$ with no edges between vertices of $D$, and one external generator for each subset of $D$ of cardinality at least $s$. The number of edges of $E_s$ is clearly

$$\beta(s) = \sum_{i \geq s} i \binom{d}{i}.$$

and so $ad(E_s) = 2\beta(s)/(\alpha(s) + d)$.

In order to define $I_s$, we first define the graph $F(D)$ with vertex set $D$ of size $d$ by taking a complete graph on $D$ and removing a maximum matching. If $d$ is even, all $(d-1)$-element subsets of $D$ are generated by vertices of $D$. If $d$ is odd, the vertices of $D$ generate all the $(d-1)$-element subsets bar one, and also the full set $D$. (Evidently, $F(D)$ maximises the sum of the cardinalities of the distinct internally generated subsets of $D$.)

For $0 \leq s \leq d-1$, the graph $I_s$ based on a set $D$ of size $d$ is defined by taking the graph $F(D)$ and adding an external generator for each subset of $D$ of size at least $s$ that is not internally generated. Thus $D \subset X(I_s)$, and $I_s$ has $\alpha(s)$ vertices. The number of edges of $I_s$ is easily calculated to be

$$\beta(s) - \frac{1}{2}d^2 \quad d \text{ even},$$
$$\beta(s) - \frac{1}{2}(d^2+1) \quad d \text{ odd}.$$

By definition, $E_s$ and $I_s$ both minimally generate $\mathcal{D}_s$. In particular, $E = E_0$ and $I = I_0$ are both $d$-shattering graphs. Now, consider, as functions of $s$,

$$\begin{aligned} ad(E_s) &= 2\frac{\beta(s)}{\alpha(s)+d}, \\ ad(I_s) &= \begin{cases} 2(\beta(s) - \frac{1}{2}d^2)/\alpha(s) & d \text{ even}, \\ 2(\beta(s) - \frac{1}{2}(d^2+1))/\alpha(s) & d \text{ odd}. \end{cases} \end{aligned}$$

Suppose that the maxima of these functions are obtained (respectively) at $s^*(E)$ and $s^*(I)$. The following result is straightforward, though the details are a little lengthy, and the proof is omitted. Note that all the quantities mentioned in this and the next result depend implicitly on $d$.

**Proposition 3** *$m(I) = ad(I_{s^*(I)})$ and $m(E) = ad(E_{s^*(E)})$.*

In the next section, we prove the following result.

**Proposition 4** *Let $s \geq \min(s^*(E), s^*(I))$ and suppose that $H$ is any graph which minimally generates $\mathcal{D}_s$. Then we have*

$$ad(H) \geq \min(ad(E_s), ad(I_s)).$$

Given the two Propositions, the problem of determining $\mu(d)$ then reduces to finding $s^*(E)$ and $s^*(I)$, and selecting which of the graphs $E_{s^*(E)}$ and $I_{s^*(I)}$ has the smallest average degree. Indeed, suppose for instance this is $I_{s^*(I)}$. Now suppose $H$ shatters $D$ and $m(H) < m(I)$. Let $H_{s^*(I)}$ be a subgraph of $H$ which minimally generates $\mathcal{D}_{s^*(I)}$. Then

$$m(H) \geq ad(H_{s^*(I)}) \geq ad(I_{s^*(I)}) = m(I),$$

which is a contradiction.

Comparing the various average degrees $ad(E_s)$ and $ad(I_s)$ is a routine exercise. The following result gives us the values of $s^*(E)$ and $s^*(I)$ for every $d$, and tells us which of $ad(E_{s^*(E)})$ and $ad(I_{s^*(I)})$ is smaller. In particular, it implies that $\min(s^*(E), s^*(I)) \geq d-3$, so that, to prove Proposition 4, we need only consider the cases $s = d-1, d-2, d-3$.

**Lemma 5** *With $E_s$ and $I_s$ as defined earlier, we have:*

$$ad(E_s) \textit{ is maximized at } s = \begin{cases} d-1 & d = 2, 3 \\ d-2 & 4 \leq d \leq 9 \\ d-3 & d \geq 10 \end{cases}$$

$$ad(I_s) \textit{ is maximized at } s = \begin{cases} d-1 & d = 2 \\ d-1, d-2 & d = 3 \\ d-2 & d \geq 4 \end{cases}$$

*Furthermore, the following relationships hold.*

$$\begin{array}{llll} ad(I_{d-1}) < ad(E_{d-1}) & d \geq 2, & & \\ ad(I_{d-2}) < ad(E_{d-2}) & d \leq 4, & ad(I_{d-2}) > ad(E_{d-2}) & d \geq 5, \\ ad(I_{d-3}) < ad(E_{d-3}) & d \leq 5, & ad(I_{d-3}) > ad(E_{d-3}) & d \geq 6. \end{array}$$

Our main result may now be stated explicitly, and indeed it follows immediately from our various results.

**Theorem 6** *Let $G(d)$ be a $d$-shattering graph with minimal maximum average degree $\mu(d) = m(G(d))$. Let $K(d)$ be the largest subgraph of $G(d)$ with average degree $\mu(d)$. Then $\mu(d)$, $G(d)$ and $K(d)$ are as listed in Table 1.*

| $d$ | $G(d)$ | $K(d)$ | $\mu(d)$ |
|---|---|---|---|
| 2 | $I(2)$ | $I_1(2)$ | $4/3$ |
| 3 | $I(3)$ | $I_1(3)$ | 2 |
| 4 | $I(4)$ | $I_2(4)$ | $40/11$ |
| $5\dots 9$ | $E(d)$ | $E_{d-2}(d)$ | $2\frac{\beta(d-2)}{\alpha(d-2)+d}$ |
| $d \geq 10$ | $E(d)$ | $E_{d-3}(d)$ | $2\frac{\beta(d-3)}{\alpha(d-3)+d}$ |

Table 1: $d$-shattering graphs of minimal maximum average degree

**Corollary 7** *For $d \geq 2$ a fixed positive integer, the function $f(n) = n^{-2/\mu(d)}$ is a threshold function for almost every $G(n,p)$ to have VC dimension at least $d$, where $\mu(d)$ is as detailed in Table 1. In particular, for $d \geq 10$, a threshold function is $f(n) = n^{-g(d)}$, where*

$$g(d) = \frac{d^3 + 11d + 6}{d^2(d^2 - 3d + 8)}.$$

**Corollary 8** *Let $\gamma$ be a constant with $0 < \gamma < g(10) = 93/650$. Suppose $g(d) > \gamma > g(d+1)$. Then almost every $G(n, n^{-\gamma})$ has VC dimension equal to $d$. In particular, the VC dimension $d$ almost surely satisfies*

$$\left\lfloor \frac{1}{\gamma} \right\rfloor + 3 \leq d \leq \left\lfloor \frac{1}{\gamma} + 3 + 3\gamma \right\rfloor.$$

# 4 Proof of Proposition 4.

Fix a natural number $d \geq 2$, and take any $s$ with $d - 1 \geq s \geq \min(s^*(E), s^*(I)) \geq d - 3$. Let $H$ be any graph minimally generating $\mathcal{D}_s$. We are to prove that the average degree of $H$ is at least the minimum of the average degrees of $E_s$ and of $I_s$.

Let $F = H[D]$ be the subgraph induced by $H$ on $D$. We form a digraph $\vec{F}$ as follows. If $x \in D \cap X$ and $x$ generates $B_x \in \mathcal{D}_s$, direct an edge of $\vec{F}$ from $x$ to each vertex of $B_x$. If all subsets of $\mathcal{D}_s$ were externally generated in $H$, we would require $\beta(s) = \sum_{i \geq s} i\binom{d}{i}$ generating edges. Note that to generate a subset of $D$ internally requires one less edge than to generate the same subset externally. It follows that the number $e(H)$ of edges of $H$ is given by

$$e(H) = \beta(s) + e(F) - e(\vec{F}) - |X \cap D|.$$

Let $e_2 = e_2(H)$ be the number of edges of $F$ used twice for the generation of subsets in $\mathcal{D}_s$; that is, those edges having a vertex of $D \cap X$ at either end. Similarly, we shall denote by $e_1$ those edges of $F$ used exactly once in generating subsets in $\mathcal{D}_s$. We seek to minimise $m(H)$, for $H$ minimally generating $\mathcal{D}_s$, so we may assume that, for each edge of $F$, at least one of its endpoints is a generator. We then have

$$e(\vec{F}) - e(F) = 2e_2 + e_1 - (e_2 + e_1) = e_2.$$

Thus if we write $a = |X \cap D|$ (and assume, as above, that all edges in $F$ are generating edges), we have

$$ad(H) = 2\frac{\beta(s) - a - e_2(H)}{\alpha(s) + d - a} \qquad (*)$$

For a given $d$, $s$ and $a$, $ad(H)$ is thus minimised by maximising $e_2(H)$.

Note that, in our supposed extremal graph $I_s$, $e_2(I_s)$ is just the number of edges of $F(D)$, which is $d(d-2)/2$ if $d$ is even, and $(d-1)^2/2$ if $d$ is odd. Our interest in $I_s$ lies in the fact that $e_2(I_s) \geq e_2(H)$, whenever $H$ is a graph minimally generating $\mathcal{D}_s$, for any $s \leq d-1$. This inequality is a special case of Lemma 9 below.

Let us first consider the case $s = d-1$. Let $D$ be a set of size $d$, and recall that $F(D)$ is defined by taking a complete graph on $D$ and removing a maximum matching $M$. For $1 \leq a \leq d$, let $e(a)$ be the maximum number of edges of a subgraph $F_a$ of $F(D)$ with vertex set $B_a \subseteq D$ of size $a$. Clearly $B_a$ is obtained by including the smallest possible number of pairs of vertices matched by $M$. Thus

$$e(a) = |e(F_a)| = \begin{cases} \binom{a}{2} & 1 \leq a \leq \lceil \frac{d}{2} \rceil, \\ \binom{a}{2} - (a - \lceil \frac{d}{2} \rceil) & \lceil \frac{d}{2} \rceil \leq a \leq d. \end{cases}$$

Now define the graph $H_a$ as follows. Start with the subgraph $F_a$ of $F(D)$ defined on $B_a$, and add in all edges of $F(D)$ between $B_a$ and $D \setminus B_a$. (In other words, remove from $F(D)$ all the edges with neither endpoint in $B_a$.) Now add an external generator for every set in $\mathcal{D}_{d-1}$ not generated by an element of $B_a$. Evidently $H_a$ minimally generates $\mathcal{D}_{d-1}$, with $|D \cap X(H_a)| = a$ and $e_2(H_a) = e(a)$. The next lemma claims, essentially, that $H_a$ is extremal.

**Lemma 9** *Let $\mathcal{G}_{d-1}$ be the set of all graphs which minimally generate $\mathcal{D}_{d-1}$. Set*

$$\eta(a) = \max\{e_2(H) : H \in \mathcal{G}_{d-1}, |X(H) \cap D| = a\}.$$

*Then*

$$\eta(a) = \begin{cases} \binom{a}{2} & 0 \le a \le \left\lceil \frac{d}{2} \right\rceil, \\ \binom{a}{2} - (a - \left\lceil \frac{d}{2} \right\rceil) & \left\lceil \frac{d}{2} \right\rceil \le a \le d. \end{cases}$$

**Proof:** The example $H = H_a$ shows that $\eta(a)$ is at least as large as given by the formula.

Let $H$ be a graph minimally generating $\mathcal{D}_{d-1}$, with $|X(H) \cap D| = a$, and consider $H[X \cap D]$. Clearly $e_2(H)$ is at most the number of edges of $H[X \cap D]$; we claim that $H[X \cap D]$ does not contain a clique of size $\lceil \frac{d}{2} \rceil + 1$. The result will then follow by Turán's Theorem, since the formula gives the maximum number of edges in a graph on $a$ vertices with no clique of size $\lceil \frac{d}{2} \rceil + 1$.

If the subset $Y$ of $X$ of size $\lceil \frac{d}{2} \rceil + 1$ induces a complete subgraph $K_{\lceil \frac{d}{2} \rceil + 1}$ in $H$, then all the $\lceil \frac{d}{2} \rceil + 1$ different subsets (of cardinality at least $d-1$) generated by these vertices would contain all the vertices of $Y$. However, the number of subsets of $D$ of size at least $d-1$ containing all the vertices of $Y$ is at most $1 + d - (\lceil \frac{d}{2} \rceil + 1) = \lfloor \frac{d}{2} \rfloor < \lceil \frac{d}{2} \rceil + 1$, which is a contradiction. □

We are now able to complete the proof of Proposition 4. Let $H$ be a fixed graph minimally generating $\mathcal{D}_s$, for some $s$ with $d - 3 \le s \le d - 1$, with a set $X$ of generators such that $|X \cap D| = a$. Recall equation $(*)$, stating that

$$\frac{1}{2} a d(H) = \frac{\beta(s) - a - e_2(H)}{\alpha(s) + d - a},$$

where $e_2(H)$ is the number of generating edges used twice. We now consider the functions

$$\psi_I(s, a) = \frac{\beta(s) - a - \binom{a}{2}}{\alpha(s) + d - a},$$

$$\psi_{II}(a) = \frac{\beta(d-1) - a - \{\binom{a}{2} - (a - \lceil \frac{d}{2} \rceil)\}}{\alpha(d-1) + d - a}.$$

It can easily be verified that $\partial^2 \psi_I / \partial a^2 < 0$ and that the second derivative of $\psi_{II}$ is negative. It follows that the functions $\psi_I, \psi_{II}$ are concave in $a$. Observe that

$\frac{1}{2}ad(E_s) = \psi_I(s,0)$ and $\frac{1}{2}ad(I_{d-1}) = \psi_{II}(d)$ and that, for $H$ as described, $\frac{1}{2}ad(H) \geq \psi_I(s,a)$.

*Case* $s = d-1$. It is quite easy to verify that for $\lceil \frac{d}{2} \rceil \leq a \leq d$,

$$\frac{1}{2}ad(H) \geq \psi_{II}(a) \geq \psi_{II}(d) = \frac{1}{2}ad(I_{d-1}).$$

For $0 \leq a \leq \lceil \frac{d}{2} \rceil$, we have

$$\frac{1}{2}ad(H) \geq \psi_I(d-1,a) \geq \psi_I(d-1,0) = \frac{1}{2}ad(E_{d-1}).$$

(This latter inequality is easy to prove; we omit the tedious details.) Thus we see that one of $E_{d-1}$ and $I_{d-1}$ is minimal here.

*Case* $s = d-2$. Now, for $d \geq 4$ and $a \leq d-2$, we have, by the concavity of $\psi_I$,

$$\psi_I(d-2,a) \geq \min(\psi_I(d-2,0), \psi_I(d-2,d-2)) = \psi_I(d-2,0) = \frac{1}{2}ad(E_{d-2}).$$

When $a = d-1$, we clearly have $e_2(H) < \binom{d-1}{2}$ and so

$$\frac{1}{2}ad(H) \geq \frac{\beta(d-2) - (d-1) - \left[\binom{d-1}{2} + 1\right]}{\alpha(d-2)+1} \geq \frac{1}{2}ad(E_{d-2})$$

for $d \geq 4$.

*Case* $s = d-3$. It is sufficient (and easy) to verify that for $d \geq 6$, we have $\psi_I(d-3,d-1) \geq \frac{1}{2}ad(E_{d-3})$. For then, if $1 \leq a \leq d-1$, we have

$$\begin{aligned} \frac{1}{2}ad(H) &\geq \psi_I(d-3,a) \\ &\geq \min(\psi_I(d-3,d-1), \psi_I(d-3,0)) \\ &\geq \frac{1}{2}ad(E_{d-3}). \end{aligned}$$

These observations complete the proof of Proposition 4.

# 5 Quickly decreasing $p(n)$

We now show that provided $d$ does not tend to infinity too rapidly, the threshold for VC dimension $d$ is still determined by the average degree of $E_{d-3}$. To do this we apply the *Janson Inequality* in the form given in [1] to the techniques on small subgraphs as developed in Chapter IV of [5]. This allows us to bound fairly precisely the VC dimension of almost every $G(n,p)$ provided $p = p(n)$ tends to 0 fast enough.

A graph $G$ is said to be *strictly balanced* if the average degree of any subgraph of $G$ is strictly less than the average degree of $G$. We note that $E_{d-3}$ is strictly balanced, and also in the terminology of Chapter IV of [5], that $(E_{d-3}, E_{d-4}, \ldots, E_1, E_0)$ is a *grading* of $E = E_0$.

Let $H$ be some unlabelled graph, and let $\{A_i : i \in I\}$ be the family of edge sets of possible copies of $H$ appearing in a random graph $G = G(n,p)$. Let $B_i$ be the event that $A_i \subset G$. Let $\mu = \sum \mathbf{Pr}(B_i)$, $\epsilon = \mathbf{Pr}(B_i)$, and $\Delta = \sum_{i\sim k} \mathbf{Pr}(B_i \wedge B_k)$, where $i \sim k$ if $i \neq k$ and $A_i \cap A_k \neq \emptyset$. The Janson Inequality (see, for instance, [1], p.96) states that

$$\mathbf{Pr}(\bigwedge_{i\in I} \overline{B}_i) \leq \exp\left(-\mu + \frac{1}{1-\epsilon}\frac{\Delta}{2}\right).$$

**Theorem 10** *Suppose $\omega = \omega(n) \to \infty$, and suppose $p = p(n)$ satisfies*

$$\exp\left(-\log n/\omega(n)\right) \leq p(n) \leq \exp\left(-\sqrt{\omega(n)\log n}\right).$$

*Then the VC dimension of a random graph $G(n,p)$ is almost surely at least*

$$l(n) = \left\lfloor \frac{\log n}{\log(1/p)} + 3 - 10\frac{\log(1/p)}{\log n} \right\rfloor$$

*and almost surely at most*

$$u(n) = \left\lfloor \frac{\log n}{\log(1/p)} + 3 + 4\frac{\log(1/p)}{\log n} \right\rfloor.$$

**Proof:** Note that with $p$ in the specified range, we have

$$\omega(n) \leq l(n) \leq u(n) = O\left(\sqrt{\frac{\log n}{\omega(n)}}\right).$$

Let $d = l(n)$. We will be mainly concerned with showing that the graph $E_{d-3}$ can almost surely be found as a subgraph in $G(n,p)$. We set $\alpha = \alpha(d-3)$ and $\beta = \beta(d-3)$, and observe that $E_{d-3}$ has $\alpha + d$ vertices, $\beta$ edges, and automorphism group of size $d!$.

Now, we may choose $\delta > 0$ such that $p = (1+\delta)n^{-g(d)}$, where, as in Theorem 6,

$$g(d) = \frac{\alpha + d}{\beta} = \frac{d^3 + 11d + 6}{d^2(d^2 - 3d + 8)},$$

and $(4d \log d)/\beta \le \delta \le 1/(\omega d^2)$.

We show that, for these values of $p(n)$ and $d$, the random graph $G(n,p)$ almost surely contains a copy of $E = E_0$ as a (not necessarily induced) subgraph. We shall then check that a copy of $E$ almost surely exists as an induced subgraph, which implies the result.

By Theorem 6, $E_{d-3}$ is the maximal average degree subgraph of $E$ and moreover it is easy to check that $E_{d-3}$ is strictly balanced. We follow the proof of Theorem 12 of Chapter IV of Bollobás [5]. Thus we partition the vertex set $V$ into subsets $V_1, \ldots, V_d$ of size $\lfloor n/d \rfloor$. Note that, as in the proof mentioned, provided $V_1$ contains a copy $H$ of $E_{d-3}$, it is straightforward to extend this copy to an $E_{d-4}$ (almost surely) using the edges from $V_2$ to $H$. Indeed, one may proceed inductively, extending $E_{d-r}$ to $E_{d-r-1}$ with edges from the vertices of $V_{r-2}$; thus extending $E_{d-3}$ to $E = E_0$ by the aforementioned grading. We omit the details.

It remains to be shown that the restriction of $G(n,p)$ to $V_1$ almost surely contains a copy of $E_{d-3}$. Let $\{A_i : i \in I\}$ be the collection of edge-sets of potential copies of $E_{d-3}$ in $V_1$, so the index set $I$ has size $(\lfloor n/d \rfloor)_{(\alpha+d)}/d!$. We bound $\Delta$ as in the proof of Theorem 1.1 of Chapter 10 of Alon and Spencer [1], and follow that notation. Let $A_i$ and $A_j$ be two distinct potential copies of $E_{d-3}$ with $m = m(i,j) = |A_i \cap A_j| \ge 1$ edges in common, and let $k = k(i,j)$ be the number of vertices spanned by $A_i \cap A_j$. Then we have $2 \le k \le \alpha + d$, and $m/k \le \beta/(\alpha + d)$ by strict balance. If $i \sim j$ then $\mathbf{Pr}(B_i \wedge B_j) = p^{|A_i \cup A_j|} = p^{2\beta - m}$ and

$$\begin{aligned} \Delta &\le (1+\delta)^{2\beta} \sum_{k=2}^{\alpha+d} \sum_{m>0} n^{2(\alpha+d)-k} n^{-((\alpha+d)/\beta)(2\beta - m)} \\ &\le e^{2\delta\beta} \sum_{k=2}^{\alpha+d} \sum_{m>0} n^{\frac{\alpha+d}{\beta}m - k}. \end{aligned}$$

We wish to bound the sum from above, so we are interested in maximising the exponent $\frac{\alpha+d}{\beta}m - k$. A little thought will show that, in order to maximise this, we should consider the case where the two copies $A_i$ and $A_j$ of $E_{d-3}$ are based on sets with $s \geq d-3$ common elements, and $t \geq 0$ external generators in common, where $s + t = k$. In the subsequent analysis we consider only the case where all the common external generators correspond to $(d-3)$-subsets, as this case contributes most to the sum, and the analysis for other cases is similar. Hence $m = t(d-3)$, and $d - 3 \leq s \leq d$. The contribution to $\Delta$ from this case is

$$\sum_t \binom{\binom{d}{3}}{t} n^{-d+3} n^{\frac{t}{\beta}((\alpha+d)(d-3)-\beta)},$$

and $(\alpha + d)(d-3) - \beta = (d^2/2)(1 + O(1/d))$. This sum is bounded above by $n^{-d/3+3}2^{d^3}$ and hence $\Delta = O(2^{d^3} n^{-d/3+3})$. Also

$$\begin{aligned} \mu &= \binom{\lfloor n/d \rfloor}{d}(n-d)_\alpha\, p^\beta \\ &\geq \frac{(1+\delta)^\beta}{d^{2d+1}} \\ &\geq d^d \text{ since } \delta \geq \frac{4d\log d}{\beta}. \end{aligned}$$

Hence, by the Janson inequality, the probability that there is no copy of $E_{d-3}$ in $V_1$ is at most

$$\exp(-d^d + O(2^{d^3} n^{-d/3+3}))$$

which tends to zero since $d = o(\sqrt{\log n})$.

We have now shown that there is almost surely a copy of $E$ as a subgraph of $G(n, p)$. We need to show that there is almost surely a copy in which there are no other edges between the base set and the external generators.

We order the potential copies of $E$ arbitrarily, and consider the event $A_k$ that the first copy of $E$ that is realised in the random graph is the $k$th, and that, in this copy, there are no such extra edges. This is the intersection of the events that (i) the edges of $E$ are all present, (ii) no earlier copy of $E$ is realised, and (iii) there are no extra edges between the base set and the external generators. Conditioned on (i), events (ii) and (iii) are both monotone decreasing events so, by Kleitman's Lemma (see, for instance, [6]), they are positively correlated. The probability of (iii), given (i), is at least $1 - pd2^d$. Hence the probability of $A_k$ is at least $1 - pd2^d$

times the probability that the $k$th potential copy of $E$ is the first to be realised. The $A_k$ are disjoint events, so the probability that one of them occurs is at least $1-pd2^d$ times the probability that some copy of $E$ is realised. We now observe that $1-pd2^d \le n^{-1/(d-4)}d2^d = 1-o(1)$, since $d = o(\sqrt{\log n})$. Hence, almost surely, one of the $A_k$ does occur, and some $d$-set is shattered.

This completes the proof of the lower bound.

We next prove that for $p(n)$ in the relevant range, almost every $G(n,p)$ has VC dimension less than $u(n)$. Now, if a graph $G$ has VC dimension at least $d$ and a subset $D$ of $d$ vertices is shattered, then $G$ contains an isomorphic copy of at least one of the at most $2^{\binom{d}{2}}$ possible graphs minimally generating $\mathcal{D}_{d-3}$. As we have seen, for $d \ge 10$, any such graph has at least $\alpha(d-3) \ge d^3/6$ vertices and average degree at least $ad(E_{d-3}) \ge 2(d-3-3/d)$. We take one such graph $H$, with $r$ vertices, and estimate the expected number of copies of $H$ in $G_{n,p}$, which is at most

$$n^r p^{ad(H)r/2} \le \left(np^{d-3-3/d}\right)^r.$$

Now set $d = \log n/\log(1/p)+3+\epsilon$, where $\epsilon \ge 4\log(1/p)/\log(n)$. Note that $\epsilon \ge 4/d$, so that $d-3-3/d \ge \log n/\log(1/p)+1/d$. Then we have

$$np^{d-3-3/d} \le p^{1/d},$$

so the expected number of copies of $H$ is at most $p^{r/d} \le p^{d^2/6}$. Therefore the expected number of copies of any graph minimally generating $\mathcal{D}_s$ is at most $2^{\binom{d}{2}}p^{d^2/6} \le (2p^{1/6})^{d^2} = o(1)$. This completes the proof. □

# References


[1] Noga Alon and Joel Spencer, *The Probabilistic Method*, Wiley, 1992.

[2] Martin Anthony and Norman Biggs, *Computational Learning Theory: An Introduction*, Cambridge University Press, Cambridge, UK, 1992.

[3] Martin Anthony, Graham Brightwell, Dave Cohen and John Shawe-Taylor, On exact specification by examples, in *COLT'92, Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, July 1992, ACM Press.

[4] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler and Manfred Warmuth, Learnability and the Vapnik-Chervonenkis dimension, *Journal of the ACM*, 36(4), 1989: 929–965.

[5] Béla Bollobás, *Random Graphs*, Academic Press, 1985.

[6] Béla Bollobás, *Combinatorics*, Cambridge University Press, 1986.

[7] David Haussler and Emo Welzl, Epsilon-nets and simplex range queries. *Discrete & Computational Geometry*, 2: 127–151, 1987.

[8] David Pollard, *Convergence of Stochastic Processes*, Springer-Verlag, New York, 1984.

[9] Kathleen Romanik and Carl Smith, Testing geometric objects, Technical Report UMIACS-TR-90-69, CS-TR-2437, University of Maryland, Maryland, 1990.

[10] V.N. Vapnik, *Estimation of Dependences Based on Empirical Data*. Springer Verlag, New York, 1982.

[11] V.N. Vapnik and A. Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2), 1971: 264-280.