Total Shatterings of Graphs*

Matthew Dalton — *University of Utah*
Laurel Paget-Seekins — *Oberlin College*
Christino Tamon† — *Clarkson University*
Doug VanNort — *SUNY Potsdam*
Brennen White — *Binghamton University*

July 9, 2001

Abstract

For a graph $G = (V, E)$, a subset $S \subseteq V$ is called *shattered* if for every $A \subseteq S$, there is a vertex $v \in V$ such that $N(v) \cap S = A$, where $N(v) = \{u \mid (u, v) \in E \text{ or } u = v\}$. The *testing dimension* of G is the largest k such that every set of k vertices in G is shattered. For $k > 0$, we let $t(k)$ be the smallest n such that there is a graph on n vertices with a testing dimension of k. In this note, we prove some structural properties of graphs with testing dimension of k, we show that they are *universal* for the class of all graphs with at most k vertices, and we prove that $2^k + k - 1 < t(k) \leq k^2 2^k (1 + o(1))$.

1 Introduction

Let X be a set and $\mathcal{F}$ be a set of subsets of X, i.e., $\mathcal{F} \subseteq 2^X$. A set $S \subseteq X$ is said to be *shattered* by $\mathcal{F}$ if for every $A \subseteq S$, there is an element $f \in \mathcal{F}$ so that $f \cap S = A$. The *Vapnik-Chernovenkis dimension* (VC) of $\mathcal{F}$ is the size of the largest set S that is shattered by $\mathcal{F}$. The notion of shattering and VC dimension are important in statistical learning theory [13]. In combinatorics of set systems (or hypergraphs), these notions are related to a concept known as the *trace* of a set system or a hypergraph [4]. In this note we will consider shatterings for graphs.

Let $G = (V, E)$ be a finite, simple, undirected graph. For a vertex $v \in V$, the neighborhood of v is the set of vertices reachable from v via paths of length at most 1 (which includes v itself), i.e., $N(v) = \{u \mid (u, v) \in E\} \cup \{v\}$. The neighborhood of a graph $N(G)$ is the collection of all neighborhoods of G, i.e., $N(G) = \{N(v) \mid v \in V\}$. The *Vapnik-Chervonenkis dimension* of a graph $G = (V, E)$ is the VC dimension of $(V, N(G))$. More specifically, a set of vertices S is *shattered* by $N(G)$ (or G if it is clear from context) if for every subset A of S, there is a vertex $v_A \in V$ such that $A = N(v_A) \cap S$, i.e., the neighbors of v_A that are in S are precisely the vertices in A. In this case, we also say that v_A (or $N(v_A)$) *shatters* A in S or that v_A *yields* the shattering of A within S. An easy but important observation is that, if $A, B \subseteq S$, $A \neq B$, v_A shatters A in S and v_B shatters B in S, then $v_A \neq v_B$.

The VC dimension of graphs was investigated earlier by Anthony, Brightwell, and Cooper [1] and by Kranakis, Krizanc, Ruf, Urrutia, and Woeginger [8]. In this note we focus on another notion of dimension for graphs that is still related to shatterings. The *testing dimension* of a graph G is the largest k so that every subset $S \subseteq V$ of k vertices in G is shattered by $N(G)$. So while the VC dimension

*Research partially supported by NSF grant DMS-9619820 and NSA grant MAD904-00-1-0079. Part of the work done while some of the authors were part of the Clarkson-Potsdam Mathematics Research Experience for Undergraduates (REU), Summer 2000.

†Contact author: tino@clarkson.edu

measures the size of *a* largest shattered subset of V, the testing dimension measures the largest size of k such that *all* subsets of size k in G is shattered. In the context of graphs, the notion of testing dimension was mentioned in [1]. We mention that testing dimension of set systems has been studied by others, notably by ROMANIK [11, 10] in relation to testing geometric objects in computational geometry and to approximate testing of concept classes in PAC learning theory (also by ANTHONY, BRIGHTWELL and SHAWE-TAYLOR [2]), and by SEROUSSI and BSHOUTY in relation to the logical testing of circuits [12].

We describe some notation that we used throughout this note. Let $t(k)$ be the smallest n such that there is a graph on n vertices with a testing dimension of k. For a graph $G = (V, E)$ and for a subset $U \subseteq V$ of vertices in G, let $G[U]$ denote the induced subgraph with vertex set U in G. The degree of a vertex v is denoted $deg(v)$. We denote $\delta(G)$ to be the minimum degree of G (smallest degree of a vertex) and $\Delta(G)$ to be the maximum degree of G (largest degree of a vertex). If G is a connected graph, then the vertex-connectivity of G, i.e., the minimum number of vertices of G whose removal disconnects G, is denoted $\kappa(G)$. A graph G is k-connected if $\kappa(G) \geq k$. We let $\alpha(G)$ denote the size of the largest independent set (empty subgraph) of G. A claw is a $K_{1,n}$, for some $n > 0$. The *center* of a claw is a vertex with maximum degree. All references to cycles and paths refer to simple cycles and paths.

2 Structure

The smallest graph with a testing dimension equals 1 is $\overline{K}_2$. In what follows we will deal only with graphs of testing dimension of at least 2. We start by stating some simple facts and lemmas about these graphs.

Fact 1 *Let $G = (V, E)$ be a graph of testing dimension equals $k \geq 2$. Then* **(i)** *G shatters every subset of ℓ vertices, for $1 \leq \ell \leq k$;* **(ii)** *If $S \subseteq V$ is a set of k vertices, then there is a subset $U \subseteq V$ such that $|U| = 2^{|S|}$ and a bijection $\varphi : 2^S \to U$ such that for every $A \subseteq S$, $N(\varphi(A)) \cap S = A$.*

The second fact above states that we can uniquely identify vertices that yield all the shatterings within a fixed subset of k vertices in a graph of testing dimension equals k. In the next theorem, we prove a *forbidden subgraph* property of graphs with testing dimension k. We refer to this property as *disjoint claw covering.*

Theorem 2 *(Disjoint Claw Covering) Let $G = (V, E)$ be a graph whose vertices can be disjointly covered with k claws, i.e., there exists a subgraph with $|V|$ vertices that is a disjoint union of claws. Then G has a testing dimension that is strictly less than k.*

Proof: Assume that G has a testing dimension of k. Let $V(G) = \bigcup_{i=1}^{k} C_i$, where each C_i is a claw ($K_{1,t}$, for some t) with center c_i. Consider the set of vertices $C = \{c_1, \ldots, c_k\}$. Note that this set is not shattered by $N(G)$, since there is no vertex whose neighborhood will yield the empty set for C. This contradiction establishes the claim. $\Box$

Next we state some very simple observations on the degree and connectivity structure of graphs with testing dimension of at least 2.

Lemma 3 *Let $G = (V, E)$ be a graph with a testing dimension of $k \geq 2$. Then* **(i)** *the diameter of G equals 2.* **(ii)** *$\delta(G) \geq 3$.* **(iii)** *G is 2-connected.*

Proof: **(i)** Two distinct vertices $u, v \in V$, $u \neq v$, are either adjacent or has a vertex $w \notin \{u, v\}$ that shatters the doubleton $\{u, v\}$ in a k-set that includes both u and v.
(ii) If there is a vertex v of degree at most 2, then its adjacent neighbors form a disjoint covering with at most 2 claws (since G has a diameter of 2).

(iii) It suffices to show that every two vertices in G belong to a common cycle. Assume that G contains two vertices u, v such that there is no cycle that contains both of them. If u and v are adjacent, i.e., $(u, v) \in E$, then there is a bipartition of V into V_u and V_v such that (u, v) is an edge cut (removal of the edge (u, v) disconnects G). Since neither u nor v can have degree 1, there are two vertices $x \in V_u \setminus \{u\}$ and $y \in V_v \setminus \{v\}$ whose minimum distance is at least 3. This is a contradiction because G has diameter 2. Now, if u and v are not adjacent, then there has to be another vertex w that is adjacent to both u and v. Consider the doubleton $\{v, w\}$. There must be a vertex $z \notin \{v, w\}$ that is only adjacent to v (to yield a shattering of $\{v\}$). Next consider the vertex that shatters the doubleton $\{z, u\}$, say a. Now we found a cycle (a, u, w, v, z, a) that contains u and v. This contradiction completes the claim. □

We generalize the observations from the previous lemma by proving slightly more general bounds on degree and connectivity of graphs with testing dimension of $k \geq 2$.

Proposition 4 *Let $G = (V, E)$ be a graph with a testing dimension of at least $k \geq 2$.* **(i)** *For every $v \in V$, $2^{k-1} - 1 \leq deg(v) \leq |V| - 2^{k-1} + 1$.* **(ii)** *The vertex connectivity of G satisfies $2^{k-2} \leq \kappa(G) \leq |V| - 2^{k-1} + 1$.*

Proof: **(i)** Consider a set of k vertices S of G and a vertex $v \in S$. There are 2^{k-1} subsets of S that contain v and since each of these subsets must be shattered by distinct elements of $N(G)$ (see Fact 1), there have to be at least $2^{k-1} - 1$ edges connected to v (v itself potentially accounts for shattering the singleton subset $\{v\}$). This gives $deg(v) \geq 2^{k-1} - 1$. But v is also *not* included in 2^{k-1} subsets of S. So it cannot be connected to more than $|V| - (2^{k-1} - 1)$ vertices in G.
(ii) Consider a set $S \subseteq V$ of k vertices in G and two vertices $u \neq v \in S$. Since there are 2^{k-2} subsets of S that contain both u and v, and each of these subsets must be shattered by a different element of $N(G)$ (see Fact 1), there are 2^{k-2} distinct vertices that connect u to v. Hence, there are at least 2^{k-2} vertex-disjoint paths connecting u to v, implying that the vertex connectivity of G is at least 2^{k-2} (by Menger's theorem). The upper bound on vertex-connectivity is obtained from an upper bound on the maximum degree of vertices in G (appealing to part **(i)**. □

By Theorem 6 (from the next section), the empty graph $\overline{K}_k$ on k vertices must exist in a graph with a testing dimension of k. We show that the size of the largest independent set of such a graph has to be strictly larger than k.

Corollary 5 *Let $G = (V, E)$ be a graph with a testing dimension of k. Then $\alpha(G) > k$.*

Proof: Let I be the maximum independent set of G, i.e., $|I| = \alpha(G)$. Note that every vertex $v \in V \setminus I$ must be connected to at least one vertex $x \in I$. Hence I induces a vertex covering of $V(G)$ via claws. Now Theorem 2 implies that $|I| > k$. □

3 Universality

A graph G is called *universal* for a family of graphs $\mathcal{H}$ if G contains all graphs $H \in \mathcal{H}$ as *induced* subgraphs. The notion of universal graphs was studied by Moon [9] (see also the survey by CHUNG and GRAHAM [6]) and has been studied extensively by others in relation to sparse graphs and small trees. In what follows, for $k > 0$, let $\mathcal{G}_k$ be the family of all graphs on at most k vertices.

Theorem 6 *Let $G = (V, E)$ be a graph with a testing dimension of $k \geq 2$. Then G is universal for $\mathcal{G}_k$.*

Proof: We prove the assertion using induction on k. The base case for $k = 2$ is trivial. Assume by induction that every graph with a testing dimension of less than or equal to $k \geq 2$ has the property that it contains all possible graphs of size at most k as induced subgraphs. Let G be a graph with a testing dimension of $k+1$. Recall that, by Fact 1, $N(G)$ shatters also all subsets of k vertices in G. Let H be a graph on $k+1$ vertices and let $V(H) = \{x\} \cup Y$ be a partition of the vertices of H into a vertex x and a set of k vertices Y. By the inductive hypothesis, the graph $H[Y]$ exists in G as an induced subgraph. We will abuse the notation and write Y to mean the set of vertices in G that induces the graph $H[Y]$.

Let z be a vertex such that $N(z) \cap Y = \emptyset$, i.e., z is not adjacent to all of the vertices of Y. Let $\tilde{Y} \subseteq Y$ be the set of vertices in Y that are adjacent to x in H. Take the subset $\{z\} \cup \tilde{Y}$ which contains at most $k+1$ vertices. Since G has a testing dimension of $k+1$, there exists a vertex s such that $N(s) \cap (\{z\} \cup Y) = \tilde{Y}$. Finally, the induced subgraph containing the vertices $\{s\} \cup Y$ in G is precisely the graph H. □

Using the above theorem, we describe some observations on the length of the shortest and longest cycles in a graph with a testing dimension of $k \geq 2$.

Corollary 7 *Let $G = (V, E)$ be a graph with a testing dimension of at least $k \geq 3$. Then the length of the shortest cycle in G, i.e., its girth, is 3.*

Proof: By Theorem 6, G must contain a 3-cycle. □

For our next lower bound on the length of the longest cycle, i.e., circumference, we use the following famous theorem of Dirac [7].

Theorem 8 *(Dirac) If G is 2-connected graph with a minimum degree greater than or equal to ℓ, then G is Hamiltonian or it has a cycle of length at least 2ℓ*

Corollary 9 *Let $G = (V, E)$ be a graph with a testing dimension of at least $k \geq 2$. Then the length of the longest cycle in G, i.e., its circumference, is at least $2^k - 2$.*

Proof: By Proposition 4, the minimum degree of G is at least $2^{k-1} - 1$, and by Lemma 3, G is 2-connected. Thus, by Theorem 8 of Dirac, G has a cycle of length at least $2(2^{k-1} - 1) \geq 2^k - 2$. □

4 Asymptotics and exact values of $t(k)$

Recall that $t(k)$ is the minimum n such that there is a graph on n vertices with a testing dimension of k. In this section, we study the asymptotic behavior of $t(k)$ as $k \to \infty$ and determine the exact value of $t(2)$.

Theorem 10 *For every $k \geq 2$, we have $2^k + k - 1 < t(k) \leq k^2 2^k (1 + o(1))$.*

NOTE: The first inequality is absolute whereas the second inequality is an asymptotic relation.

Proof: Let G be a minimal graph of a testing dimension of k. The lower bound follows from observing the shatterings around a complete subgraph K_k, which, by Theorem 6, we know must exist as an induced subgraph of G. Let S be the set of k vertices in G forming the k-clique. Note that all $2^k - 1$ subsets, except for the full subset S itself, must be shattered by external vertices. Hence there has to be at least $k + 2^k - 1$ vertices in G.

The upper bound is obtained by employing a basic probabilistic method [3]. We examine the random graph $G = G(n, \frac{1}{2})$. Fix a set S of k vertices in G. Let $A \subseteq S$ be a fixed subset of S. For $v \in V(G)$, let β_v be the event that v does not shatter A in S. So the probability that A is not shattered in S is given by

$$\begin{aligned}
\Pr[\bigcap_v \beta_v] &= \Pr[\bigcap_v \beta_v \mid \bigcap_{v \in S} \beta_v] \Pr[\bigcap_{v \in S} \beta_v] \\
&\le \Pr[\bigcap_v \beta_v \mid \bigcap_{v \in S} \beta_v], \quad \text{since } \Pr[\textstyle\bigcap_{v \in S} \beta_v] \le 1 \\
&\le \Pr[\bigcap_{v \notin S} \beta_v \mid \bigcap_{v \in S} \beta_v] \le \Pr[\bigcap_{v \notin S} \beta_v], \quad \text{by independence} \\
&\le \prod_{v \notin S} \Pr[\beta_v] \le (1 - 2^{-k})^{n-k}.
\end{aligned}$$

Thus, the probability that there is a subset A of S that is not shattered by $N(G)$ is at most $\binom{n}{k} 2^k (1 - 2^{-k})^{n-k} < 1$, by our choice of $n = k^2 2^k (1 + o(1))$. This completes the proof. □

In what follows, we will provide precise information about minimal graphs with testing dimension of 2.

Lemma 11 *Let $G = (V, E)$ be a graph with a testing dimension of 2. Then G contains a ℓ-cycle, for $\ell \geq 6$.*

Proof: By Lemma 3, $\delta(G) \geq 3$ and G is 2-connected. Using Theorem 8, we conclude that G has a 6-cycle or longer. □

Next we will prove that the minimum number of vertices and edges necessary and sufficient to obtain a graph with a testing dimension equals 2 are 8 and 12, respectively. We refer to a graph with the minimum number of vertices and edges as a *minimal* graph of that testing dimension.

Theorem 12 $t(2) = 8$

Proof: The upper bound is obtained by observing the following cubic graphs on 8 vertices given in Figure 1. As described in [5], there are 5 non-isomorphic cubic graphs on 8 vertices (all being Hamiltonian) and only two of them have testing dimension equals 2. See Figure 3 at the end of this note.

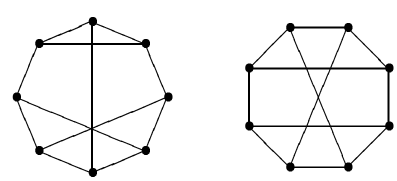

Figure 1: The (only) two minimal graphs with testing dimension 2.

The lower bound is proven by a *cycle elimination* argument. Note that by Lemma 11, G must have a cycle of length at least 6. So the number of vertices of G must be at least 6. But if $n = 6$ then a 6-cyle will induce a disjoint covering by 2 claws. So G cannot have 6 vertices.

We now consider the case where G has $n = 7$ vertices. We proceed by showing that there are no cycles of lengths $\ell = 6, 7$ in G. Suppose that there is a 6-cycle in G. Consider the first graph in Figure 2. Vertex 0 has to be connected to some vertex on the 6-cycle. Without loss of generality, assume that the edge $(0, 1)$ exists. Thus, again we obtain a vertex covering with 2 claws. So a 6-cycle cannot exist.

Figure 2: Cycles of lengths 6 and 7, respectively, in a graph on 7 vertices

Suppose that there is a 7-cycle in G. Consider the second graph in Figure 2. Since the minimum degree is at least 3 by Lemma 3, vertex 2 has to be connected to another vertex. If it is connected to either vertex 4 or 7 then the two claws centered at vertices 2 and 5 cover G. Hence vertex 2 can only be connected to either vertex 5 or 6. Without loss of generality, assume that it is connected to vertex 5. Now, vertex 7 has to be connected to either vertex $2, 3, 4$ or 5 (again by the minimum degree requirement of Lemma 3. If it is connected to vertex 2 or 5, then we get two claws centered at vertices 2 and 5. If it is connected to vertex 3, then we get two claws centered at vertices 5 and 7. Finally, if it is connected to vertex 4, then we get two claws centered at vertices 2 and 7. So in all cases, we obtain a contradiction by means of the existence of two disjoint claw covering of G. Thus, there cannot exist a 7-cycle in G. This completes the proof. $\square$

It is clear that 12 is the minimum number of edges for graphs on 8 vertices to have testing dimension 2 (since the minimum degree of a vertex must be at least 3). We have been unable to obtain precise information about graphs with higher testing dimensions, say $k = 3, 4, \ldots$.

5 Conclusions

We conclude this note with the following open questions and conjectures. **(i)** For every $k \geq 1$, there is a *regular*, *Hamiltonian* graph with the minimum number of vertices *and* edges that achieves a testing dimension of k. **(ii)** Can we show $t(k) \in \Omega(k2^k)$? Also, what is the correct asymptotic expression for $t(k)$? **(iii)** Is $16 \leq t(3) \leq 24$? **(iv)** Are there *explicit constructions* for graphs with arbitrary testing dimensions? **(v)** What is the complexity of computing the testing dimension of graphs?

Acknowledgments

We thank Travis Schedler for a comment on Theorem 10, David Powers for helpful comments, valuable pointers to literature, and providing us with [5] (which showed that there are only *two* minimal graphs with testing dimension 2, the second and third graphs in Figure 3, saving us endless time sieving through computer-generated output), and Kazem Mahdavi, Siew Hwee Lee for proof-reading the manuscript, and Amarda Shehu for pointing out an error in an earlier version of this note.

References

[1] Martin Anthony, Graham Brightwell, Colin Cooper. The Vapnik-Chervonenkis Dimension of a Random Graph. *Discrete Mathematics*, **138**:43-56, 1995.

[2] Martin Anthony, Graham Brightwell, John Shawe-Taylor. On Specifying Boolean Functions by Labelled Examples. *Discrete Applied Mathematics*, **61**:1-25, 1995.

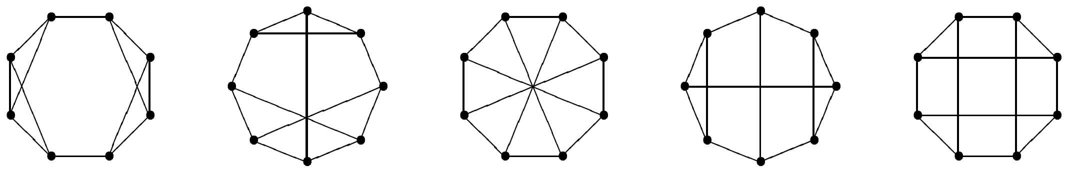

Figure 3: All cubic graphs on 8 vertices (source [5]).

[3] NOGA ALON, JOEL H. SPENCER, PAUL ERDŐS. *The Probabilistic Method.* John Wiley & Sons, 1992.

[4] BÉLA BOLLOBÁS. *Combinatorics.* Cambridge University Press, 1986.

[5] F.C. BUSSEMAKER, S. ČOBELJIĆ, D.M. CVETKOVIĆ, J.J. SEIDEL. Computer investigation of cubic graphs. Technical Report T.H.-Report 76-WSK-01, Department of Mathematics, Technological University Eindhoven, The Netherlands, January 1976.

[6] F.R.K. CHUNG, R.L. GRAHAM. On Universal Graphs. *Annals of the New York Academy of Sciences*, **319**:136-140, 1979.

[7] G.A. DIRAC. Some theorems on abstract graphs. *Proc. London Mathematical Society*, **2**, 69-81, 1952.

[8] EVANGELOS KRANAKIS, DANNY KRIZANC, BERTHOLD RUF, JORGE URRUTIA, GERHARD J. WOEGINGER. VC-dimension for Set Systems defined for Graphs. *Discrete Applied Mathematics*, **77**:237-257, 1997.

[9] J.W. MOON. On minimal n-universal graphs. *Proc. Glasgow Math. Soc.*, **7**:32-33, 1965.

[10] KATHLEEN ROMANIK. Approximate Testing and Learnability. *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, 327-332, 1992.

[11] K. ROMANIK, C. SMITH. Testing Geometric Objects. *Computational Geometry: Theory and Applications*, **4**:157-176, 1994.

[12] GADIEL SEROUSSI, NADER H. BSHOUTY. Vector Sets for Exhaustive Testing of Logic Circuits. *IEEE Transactions on Information Theory*, **34**(3):513-522, May 1988.

[13] VLADIMIR N. VAPNIK. *The Nature of Statistical Learning Theory.* Springer-Verlag, 1999.