Minimal Saturated Subgraphs of the Hypercube

Kavish Gandhi

under the direction of
Mr. Chiheon Kim
Massachusetts Institute of Technology

Research Science Institute
July 30, 2014

Abstract

Within the hypercube Q_n, we investigate bounds on the *saturation number* of a forbidden graph G, defined as the minimum number of edges in a subgraph H of Q_n that is both G-free and has the property that the addition of any $e \in E(Q_n) \setminus E(H)$ creates G. For all graphs G, we find a lower bound based on the minimum degree of non-leaves. For upper bounds, we first examine general graphs and derive conditions that, if satisfied, allow us to bound the saturation number. We also study specific cases, finding improved bounds for paths, stars, and most caterpillars. In all of these cases, we find bounds that are $O(2^n)$, an interesting fact that we conjecture to hold for all trees T.

Summary

We study properties of the hypercube, an extension of the cube to any number of dimensions. The hypercube is especially interesting because a direct connection can be made between it and parallel computer networks. Our problem, in a parallel network, is equivalent to finding the maximum number of links that must fail before some desired configuration of processors and links no longer exists. This value is directly associated with the stability of the parallel architecture. We provide improved bounds on this maximum for many desired structures, along with devising novel methods for finding both lower and upper bounds.

1 Introduction

In 1941, Paul Turán [1] proved his seminal theorem, aptly named Turán's Theorem, explicitly determining the maximum number of edges in a K_{r+1}-free subgraph of the complete graph K_n. This result sparked the study of what is now known as the extremal function. In particular, the extremal function $\mathrm{ex}(H, G)$ is defined as the maximum number of edges in a subgraph of some host graph H that does not contain some forbidden graph G. An example of the extremal function is shown in Figure 1, which shows how deleting two edges from the complete graph K_4 can create a graph free of the three-cycle C_3. It can be easily seen that deleting just one edge always maintains some three-cycle, so $\mathrm{ex}(K_4, C_3) = 4$.

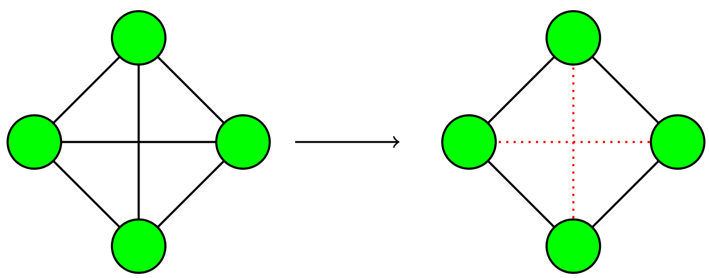

Figure 1: A simple example with $H = K_4$ and $G = C_3$. We see that $\mathrm{ex}(K_4, C_3) = 4$.

The study of $\mathrm{ex}(H, G)$ has a long history. For many years, the major focus in the field was on determining $\mathrm{ex}(K_n, G)$ and $\mathrm{ex}(K_{m,n}, G)$, which have, for the most part, been either found exactly or within constant-order bounds. We, on the other hand, study subgraphs of Q_n, the n-dimensional hypercube with vertex set $\{0, 1\}^n$ and edge set consisting of all pairs of vertices differing at exactly one coordinate. Erdös [2] was among the first to study $\mathrm{ex}(Q_n, G)$, specifically trying to determine $\mathrm{ex}(Q_n, C_4)$. This question is still open today; the best bounds, due to Bialostocki and Balogh et al. [3, 4], are $(n + \sqrt{n})2^{n-2} \leq \mathrm{ex}(Q_n, C_4) \leq 0.6068n2^{n-1}$. Erdös also conjectured that, for all $k \geq 2$, $\mathrm{ex}(Q_n, C_{2k}) = o(e(Q_n))$, but was proven wrong by Chung [5], who showed that $\mathrm{ex}(Q_n, C_6) \geq \frac{1}{4}e(Q_n)$. However, she also showed that $\mathrm{ex}(Q_n, C_{4t}) = o(e(Q_n))$ for $t \geq 2$. These results were expanded by Füredi and Özkahya

[6], who proved that $\mathrm{ex}(Q_n, C_{4t+2}) = o(e(Q_n))$ for $t \geq 3$, a result which was later shown in a more general framework by Conlon [7]. The only remaining unresolved asymptotic case is C_{10}, which, despite some progress, remains open.

The extremal function also has a natural opposite formulation. In particular, consider the following definition:

Definition 1.1. A subgraph H' of H is G-*saturated* if it is G-free, but the addition of any edge in $E(H) \setminus E(H')$ creates a copy of G.

The extremal function defines the maximum number of edges in such a saturated graph, but we can also ask: what is the minimum number of edges in a saturated graph? To this end, we define the *saturation number* $\mathrm{sat}(H, G)$ as the minimum number of edges in a G-saturated subgraph of H. An example of saturation is shown in Figure 2, in which the addition of any non-edge to the red subgraph creates C_4, though the original subgraph does not contain C_4. Thus, the original red subgraph is saturated, and, as it turns out, this subgraph has the minimum number of edges possible; in other words, $\mathrm{sat}(Q_3, C_4) = 8$.

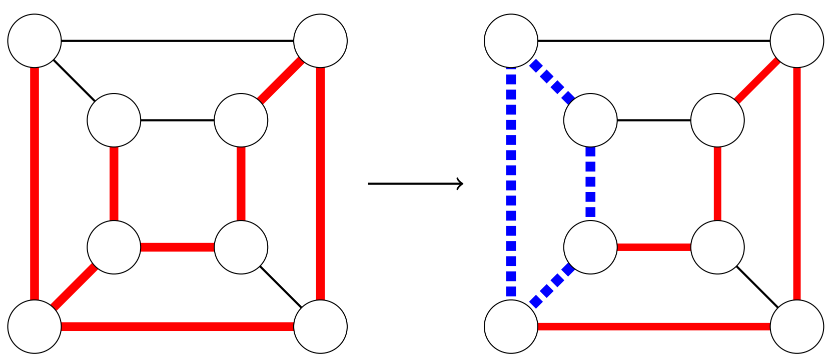

Figure 2: An example of a C_4-saturated subgraph of the cube. Shown on right is the C_4 created by the addition of a blue dotted non-edge.

However, though $\mathrm{ex}(Q_n, G)$ and in particular $\mathrm{ex}(Q_n, C_{2k})$ have been very well studied, $\mathrm{sat}(Q_n, G)$ has not. The best early result was of Choi and Guan [8], who showed that

$$\lim_{n\to\infty} \frac{\mathrm{sat}(Q_n, Q_2)}{e(Q_n)} \leq \frac{1}{4}.$$

Very recently, Johnson and Pinto [9] improved this result, showing that

$$\lim_{n\to\infty} \frac{\mathrm{sat}(Q_n, Q_m)}{e(Q_n)} = 0.$$

To do this, they proved Theorem 1.2 using an explicit inductive construction:

Theorem 1.2 (Johnson and Pinto, 2014). *For all $m \geq 1$, there exist constants c_m and a_m such that* $\mathrm{sat}(Q_n, Q_m) \leq \frac{c_m}{n^{a_m}} \cdot e(Q_n)$. *More precisely, $a_1 = 1$, and $a_m = \frac{1}{7\cdot 3^{m-2}}$ for all $m > 1$.*

Johnson and Pinto also established explicit bounds for $\mathrm{sat}(Q_n, Q_2)$, finding that

$$\frac{3}{2} \cdot 2^n \leq \mathrm{sat}(Q_n, Q_2) < 10 \cdot 2^n.$$

These results were important in opening the study of $\mathrm{sat}(Q_n, G)$. However, little work has been done in determining the order of $\mathrm{sat}(Q_n, T_k)$ for trees with k edges. We examine this problem, both for general trees and for some specific cases. For general trees, we find several bounds, showing that for trees decomposable into subtrees with smaller cubical dimension and for trees with sufficiently large minimum degree with respect to their diameter, $\mathrm{sat}(Q_n, T_k) \leq c \cdot 2^n$ for some constant c that depends on k. We also find several improved upper bounds for specific types of trees. Finally, for all trees, we find lower bounds on the order of 2^n, with the constant factor related to the minimum degree of the graph.

Now, we present an outline of our paper. In Section 2, we define additional terms that are used commonly in our paper. Then, in Section 3, we prove some general theorems and lemmas that establish a lower bound for the saturation number of general graphs and provide the foundation for some of our explicit upper bounds for trees. In Section 4, we examine the saturation number of trees. We first show a general upper bound for trees that satisfy a particular condition, and then determine some classes of trees that have this property. Then, we find improved upper and lower bounds for paths of length k, stars, generalized stars, and

k-stars. Finally, in Sections 5 and 6, we summarize our work, its contribution to the study of saturation numbers in the hypercube, and acknowledge those involved in the writing and editing of the paper, along with the research itself.

2 Definitions

In this section, we provide definitions for some standard terms used frequently in our paper.

Definition 2.1. The *weight* of a vertex $v \in V(Q_n)$, denoted by $w(v)$, is the number of 1's in the coordinate representation of v.

Definition 2.2. The *Cartesian product* $G \,\square\, H$ is the graph created by placing copies of G at all of the vertices of H and connecting corresponding vertices of adjacent G's in H.

We specifically use the fact that $Q_n = Q_k \,\square\, Q_{n-k}$ to create some of our saturated subgraphs.

Definition 2.3. A *tree* is any connected, acyclic graph. A *leaf* of a tree is any vertex with degree one. We will denote the tree with k edges by T_k.

Now, we define some specific types of trees.

Definition 2.4. A *path* of length k is a sequence of vertices $v_1 v_2 \dots v_{k+1}$ connected consecutively by edges such that $v_i \neq v_j$ for $i \neq j$. We denote such a path by P_k.

For example, P_4 is shown in Figure 3.

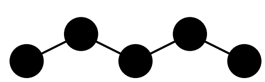

Figure 3: P_4, the path with 4 edges.

Definition 2.5. A *star* S_k is the complete bipartite graph $K_{1,k}$. In other words, S_k has one internal vertex and k leaves connected to this vertex.

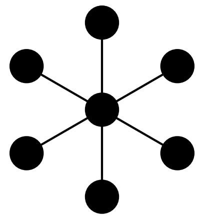

Figure 4: S_6, the star with 6 edges.

S_6 is shown as an example in Figure 4.

Definition 2.6. A *generalized star* $GS_{k,m}$ consists of one internal vertex and k disjoint paths of length m emanating from this vertex. We call each disjoint path a *leg* of $GS_{k,m}$.

For example, $GS_{3,2}$ is shown in Figure 5.

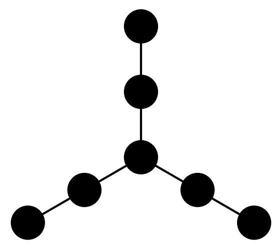

Figure 5: $GS_{3,2}$, the generalized star with 3 legs and 2 edges per leg.

Definition 2.7. A $m_1 \times m_2 \times \cdots \times m_k$ *k-star* is the tree made up of stars S_{m_1}, S_{m_2}, ..., S_{m_k} all connected in sequence by their central nodes. We denote this by $S_{m_1 \times m_2 \times \cdots \times m_k}$.

An example of a 2-star is shown in Figure 6, specifically $S_{3\times 5}$.

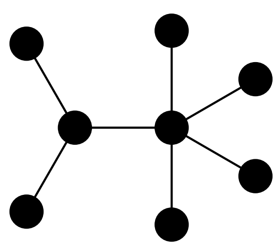

Figure 6: $S_{3\times 5}$, the 2-star consisting of a connected S_3 and S_5.

A k-star is more commonly known as a *caterpillar*, which is traditionally defined as a tree in which all vertices either lie on or are adjacent to the central path. We use these two terms interchangeably where convenient; in particular, we often use k-star to describe caterpillars with large average degree.

Our final definition is a special subset of the vertices of the hypercube that we use in Section 4.4 as a basis for our construction of saturated subgraphs.

Definition 2.8. Given $n = 2^i - 1$, let H be an i by n matrix whose columns are the nonzero vectors in $\mathbb{F}_2^i$. Then the *hamming code* on Q_n is the nullspace of H over $\mathbb{F}_2$.

A hamming code C satisfies three essential properties, all derivable from its definition:

1. $|C| = \frac{2^n}{n+1}$.

2. The distance between all pairs of vertices in C is at least 3.

3. C dominates Q_n. In other words, every vertex in $V(Q_n) \setminus C$ is adjacent to exactly one vertex in C.

An example of a hamming code in Q_3 is shown in Figure 7.

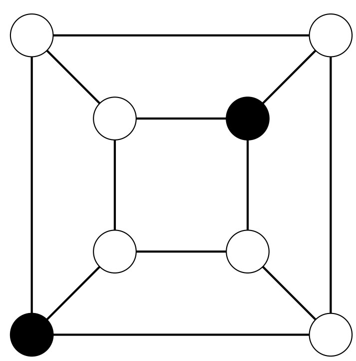

Figure 7: A hamming code, represented by the black nodes, in Q_3.

3 General Bounds and Methods

In this section, we present two preliminary results that will become important in the proofs of later results.

Our first result concerns lower bounds on $\operatorname{sat}(Q_n, G)$. We use an argument based on the minimum degree of G to find a lower bound, which is best for graphs with large minimum degree. Before stating the theorem, it is necessary to first define $\operatorname{emin}(G)$ as the minimum value, over all pairs of adjacent vertices in G, of the maximum degree of two adjacent vertices.

Theorem 3.1. *Given some graph G with $\operatorname{emin}(G) = \delta + 1$, $\operatorname{sat}(Q_n, G) \geq \delta \cdot 2^{n-3}$.*

Proof. The details of this proof can be found in Appendix A. □

Our second and final preliminary result sets the foundation for many of our inductive constructions in Section 4. In essence, it allows us to classify exactly when we can scale a saturated graph in Q_i up to Q_n while still maintaining its saturation.

Before we state this lemma, however, we first need to make an important definition.

Definition 3.2. In a G-saturated subgraph of Q_k, an *endpoint* is a vertex in Q_k such that, if an edge were to be added incident to this vertex, some isomorphism of G would be formed.

Lemma 3.3. *Given a graph G, if there exists a positive integer k such that we can construct a G-saturated subgraph of Q_k with $C \cdot e(Q_k)$ edges in which at least half of all vertices in Q_k are endpoints and there exists a cyclic rotation of these endpoints into the remaining vertices in Q_k, then $\operatorname{sat}(Q_n, G) \leq C \cdot k2^{n-1}$.*

Proof. Let H_0 be such a saturated subgraph of Q_k. By the definition of H_0, we can construct an isomorphism of H_0 in Q_k, say H_1, which contains as endpoints the remaining vertices in Q_k. This implies that any edge constructed between two Q_k's containing H_0 and H_1, respectively, must be incident to an endpoint, thereby creating a copy of G. Furthermore, H_0 and H_1 are both themselves saturated, so the larger subgraph is evidently G-free. Therefore, this new subgraph of Q_{k+1} is G-saturated. Now, if we instead perform this process in a subgraph of $Q_k \,\square\, Q_{n-k}$ in which vertices in Q_{n-k} with even weight contain H_0, vertices with odd weight contain H_1, and no edges connect adjacent Q_k's, then we have a subgraph of Q_n

in which any edge added creates G, but the initial subgraph is G-free. From this construction, we get that $\mathrm{sat}(Q_n, G) \leq |H_0| \cdot 2^{n-k} = C \cdot k2^{n-1}$, so we are done. $\square$

4 Bounds on $\mathrm{sat}(Q_n, T_k)$

In this section, we study the saturation number of trees. We begin by examining bounds for general trees in Section 4.1, and then derive tighter bounds for some special cases in Sections 4.2 and 4.3. From these cases, then, in Section 4.4, we develop a method based on the hamming code to find upper bounds for trees with large minimum degree.

4.1 The General Case

For all trees T, we define the *cubical dimension* of T, denoted by $\mathrm{cd}(T)$, as the smallest positive integer such that T can be embedded in $Q_{\mathrm{cd}(T)}$. This leads us to the following theorem, which establishes a sufficient condition to show that $\mathrm{sat}(Q_n, T) \leq c \cdot 2^n$ for some constant c.

Theorem 4.1. *Let*

$$k := \min_{e \in T} \{\max(\mathrm{cd}(T_1), \mathrm{cd}(T_2))\},$$

where T_1 and T_2 are the two subgraphs of T created by the removal of some edge e. Then, if $k < \mathrm{cd}(T)$, there exists a T-saturated subgraph of Q_n with $k \cdot 2^{n-1}$ edges.

Proof. Without loss of generality, let $\mathrm{cd}(T_1) = k$ and $\mathrm{cd}(T_2) = j$ for some $j \leq k$. Now, consider a subgraph of $Q_k \,\square\, Q_{n-k}$. We allow in this subgraph only edges within each Q_k, and no edges between Q_k's. Because $\mathrm{cd}(T) > k$, we see that this subgraph is T-free. Now, consider the new graph created when any edge is added between two Q_k's. Due to rotational symmetry, we can find any isomorphism of T_1 and T_2 in each of the Q_k's. Because T is constructed by adding an edge between T_1 and T_2, the addition of any non-edge necessarily creates T. This means that our subgraph is T-saturated, so it remains to count the number of edges, which is just $2^{n-k} \cdot k \cdot 2^{k-1} = k \cdot 2^{n-1}$. Thus, we are done. $\square$

Notice that a theorem similar to Theorem 4.1 also holds for graphs with a bridge in them. That is, if we define $B(G)$ as the set of edges in a graph such that if $e \in B(G)$ is removed, two disjoint connected components are created, then the same conclusion follows.

4.2 Paths of length k

We now provide lower and upper bounds on the saturation number of paths in the hypercube.

Proposition 4.2. *Given a path P_k of length $k \geq 2$, $\mathrm{sat}(Q_n, P_k) \geq 2^{n-3}$ for all k.*

Proof. Note that $\mathrm{emin}(P_k) = 2$, so our lower bound follows from Theorem 3.1. □

Before presenting our upper bound on the saturation number of paths, it is first necessary to prove an important lemma on the maximum length of a path in Q_k after the deletion of some set of vertices of the same parity.

Lemma 4.3. *The maximum length of a path in a subgraph of Q_k created by deleting all edges incident to j vertices of the same parity is $2^k - 2j$, where $k \geq 2$ and $0 < j \leq 2^{k-1}$.*

Proof. We prove this by induction on j and k. Begin with the base case $k = 2$, in which it is clear that deleting one vertex limits the path length to exactly $2^2 - 2(1) = 2$, and deleting two vertices of the same parity deletes all paths. Now, we induct on k. Without loss of generality assume we are deleting vertices of even weight. Given that in Q_{k-1} deleting j vertices of even weight makes the maximum path length $2^{k-1} - 2j$, we want to show that the same applies in Q_k. To do this, first split Q_k into two copies of Q_{k-1}, $(0, \{0,1\}^{k-1})$ and $(1, \{0,1\}^{k-1})$, and delete j vertices of even weight from the latter. We thus know that the longest path within $(1, \{0,1\}^{k-1})$ is $2^{k-1} - 2j$, and that the longest path within $(0, \{0,1\}^{k-1})$ is simply $2^{k-1} - 1$, created by deleting some edge from a Hamiltonian cycle in $(0, \{0,1\}^{k-1})$. Connecting the two, which is possible because vertices of odd weight in $(1, \{0,1\}^{k-1})$ still have edges incident in the first direction connected to them after vertex deletion, a path of length $2^k - 2j$ is formed. All that remains now is to show that there is no longer path. To do this, first note that there are $2^{k-1} - j$ vertices of even weight. Now, every path clearly must consist of vertices of alternating parity, so, assuming that we start and end our path with an vertex of odd parity, there are at maximum $2^{k-1} - j + 1 + 2^{k-1} - j = 2^{k-1} - 2j + 1$ vertices in this path. Any other pairing of endpoint parities leads to a smaller maximum number of vertices, so $2^{k-1} - 2j + 1$ is our global maximum. From there, since $2^{k-1} - 2j + 1$ vertices corresponds to a maximum path length of $2^{k-1} - 2j$, we are done.

With the induction on k finished, we induct on j. Having already shown the base case Q_2, assume that our result is true for $j \leq 2^{i-1}$. From this, we wish to show the same result for $2^{i-1} < j \leq 2^i$. To do this, consider Q_{i+1}, and split it into two Q_i's, $(0, \{0,1\}^i)$ and $(1, \{0,1\}^i)$. In $(1, \{0,1\}^i)$, we delete all 2^{i-1} vertices of even weight, while in $(0, \{0,1\}^i)$ we delete $j - 2^{i-1}$ vertices. Clearly, in this construction, all paths are eliminated in $(1, \{0,1\}^i)$, while, in $(0, \{0,1\}^i)$, there is a maximal path of length $2^i - 2(j - 2^{i-1}) = 2^{i+1} - 2j$. Noting that this path in $(0, \{0,1\}^i)$ must begin and end with vertices of odd weight and that their adjacent vertices in $(1, \{0,1\}^i)$ are of even weight and are therefore eliminated, this is maximal, so we are done. $\square$

Theorem 4.4. *Given $k = 2^i + j$ for $0 < j \leq 2^i$, $\operatorname{sat}(Q_n, P_k) \leq E$, where*

$$E = \begin{cases} (i+1)2^{n-2} + (i+1)(j-1) \cdot 2^{n-i-2} & \text{if } j \equiv 1 \pmod 2 \\ (i+2)2^{n-2} + i(j-1) \cdot 2^{n-i-2} & \text{if } j \equiv 0 \pmod 2. \end{cases}$$

Proof. The proof of this relies on a construction similar to that described in Lemma 4.3. Due to space constraints, it is left to Appendix B. $\square$

4.3 Saturation Number of Stars

In this section, we derive lower and upper bounds for the saturation number of various types of stars, including regular stars, generalized stars, and k-stars.

4.3.1 Regular Stars

We start by studying just a single star with k edges.

Theorem 4.5. *Given the star S_k, $\mathrm{sat}(Q_n, S_k) \geq (k - 1 + o(1))\, 2^{n-2}$ for all $k \geq 2$.*

Proof. Here, we can use Theorem 3.1 to get a lower bound of $(k-1) \cdot 2^{n-3}$. However, stars are special in that *every* non-edge must have some incident vertex with degree $k-1$. Using this, we can improve our lower bound. In particular, consider some graph H which is S_k-saturated, and define V_i as the number of vertices v in H with $\deg(v) = i$. Note first that any vertex x with degree $i < k-1$ in H must have $n-i$ neighbors that are not in H, all of which must have degree $k-1$ to maintain saturation. From here, consider the number of vertices with degree $k-1$ counted for every such vertex x,

$$\sum_{i=0}^{k-2} (n-i)V_i. \tag{1}$$

Since every vertex with degree $k-1$ is counted at most $n-k-1$ times in (1), we have that

$$\sum_{i=0}^{k-2} (n-i)V_i \leq (n-k+1)V_{k-1}. \tag{2}$$

At this point, it is also useful to note that $\sum_{i=0}^{k-1} iV_i = 2e(H)$, as each edge in H is counted twice in this summation. This leads us to the following expression as a manipulation of (2):

$$\sum_{i=0}^{k-2} (n-i)V_i = n\sum_{i=0}^{k-2} V_i - \sum_{i=0}^{k-2} iV_i. \tag{3}$$

From here, we can use our already determined bounds to simplify (3) to

$$n(2^n - V_{k-1}) - (2e(H) - (k-1)V_{k-1}) \leq (n-k+1)V_{k-1}.$$

Grouping terms, we have that

$$n2^n \leq 2e(H) + 2(n-k+1)V_{k-1}. \tag{4}$$

From here, we need a relationship between $e(H)$ and V_{k-1}. In this case, we have that

$$2e(H) \geq \sum_{i=0}^{k-1} iV_i \geq (k-1)V_{k-1},$$

so $V_{k-1} \leq \frac{2e(H)}{k-1}$. This in turn simplifies (4) to

$$n2^n \leq \frac{4(n-k+1)}{k-1}e(H),$$

or, collecting terms,

$$e(H) \geq \frac{k-1}{2n-k+1}n2^{n-1} \sim (k-1+o(1))2^{n-2}. \qquad \square$$

Theorem 4.6. *Given the star S_k,* $\text{sat}(Q_n, S_k) \leq (k - \frac{5}{3} + o_k(1))2^{n-1}$.

Proof. The proof of this can be found in Appendix C; in essence, our result improves upon the bound given by Theorem 4.1, $(k-1) \cdot 2^{n-1}$. $\square$

4.3.2 Generalized Stars

Next, we find upper bounds on the saturation number of generalized stars. Specifically, we find improved bounds for $GS_{k,2}$, $GS_{k,3}$, and $GS_{k,5}$, generalized stars with any number of legs and 2, 3, and 5 edges in each leg, respectively. We then use Theorem 4.1 to resolve some of the other cases by packing disjoint one-to-many paths of maximal length in the hypercube.

Our first two results concern upper bounds on $\mathrm{sat}(Q_n, GS_{k,2})$, $\mathrm{sat}(Q_n, GS_{k,3})$, and $\mathrm{sat}(Q_n, GS_{k,5})$. The proofs, due to space constraints, are left to Appendix D.

Theorem 4.7. *For all* $k \geq 3$, $\mathrm{sat}(Q_n, GS_{k,2}) \leq (k+1) \cdot 2^{n-2}$.

Theorem 4.8. *For all* $k \geq 4$, $\mathrm{sat}(Q_n, GS_{k,3}) \leq (k+2) \cdot 2^{n-2}$.

Theorem 4.9. *For all* $k \geq 6$, $\mathrm{sat}(Q_n, GS_{k,5}) \leq (k+2) \cdot 2^{n-2}$.

Lemma 4.10. *Given the hypercube* Q_k *and some vertex* $v \in V(Q_k)$*, there exist* k *vertex-disjoint paths of length* $k-1$ *starting at* v*.*

Proof. Let us denote the directions in the k-dimensional hypercube by $1, 2, \ldots, k$. We characterize each path in Q_m of length m by a m-tuple $(a_1, a_2, \ldots, a_m)$, where $a_i \in [k]$, representing the order of directions travelled.

Consider k paths starting at v, $\mathcal{P}_1 = (1, 2, \ldots, k-1)$, $\mathcal{P}_2 = (2, 3, \ldots, k)$, $\ldots$, $\mathcal{P}_k = (k, 1, \ldots, k-2)$. Refer to the i-tuple corresponding to the first i directions of such a path by $\mathcal{P}_{ji}$. Now, note that, in these paths, $\mathcal{P}_{ai} \neq \mathcal{P}_{bi}$ for all $a \neq b$. This implies that the paths must be vertex-disjoint, and, since they are all of length $k-1$, we are done. $\square$

Note that we suspect that the length of these k paths can actually be $O(\frac{2^k}{k})$. In particular, we believe that these paths may always be able to achieve the maximum $\left\lfloor \frac{2^k-1}{k} \right\rfloor$, which we have been able to show for $k = 2, 3, 4, 5, 6, 7$. However, we have not been able to generalize this, so we leave it as a conjecture:

Conjecture 4.11. *Given the hypercube* Q_k *and some vertex* $v \in V(Q_k)$*, there exist* k *vertex-disjoint paths of length* $\left\lfloor \frac{2^k-1}{k} \right\rfloor$ *starting at* v*.*

From Lemma 4.10, we are able to derive an upper bound on the saturation number of certain generalized stars.

Theorem 4.12. *For all positive integers k, m where $m < k-1$, $\mathrm{sat}(Q_n, GS_{k,m}) \le (k-1)\cdot 2^{n-1}$.*

Proof. To show this, we invoke Theorem 4.1. First note that our original tree $GS_{k,m}$ cannot be embedded in Q_{k-1}, as it contains a vertex with degree k. Now, delete some edge incident to the central vertex of our generalized star, thereby splitting the graph into $T_1 = P_{k-1}$ and $T_2 = GS_{k-1,m}$. Clearly, T_1 can be embedded in Q_{k-1}, and, since $m < k-1$, T_2 can also be embedded in Q_{k-1} as a direct result of Lemma 4.10. Therefore, as we now have a splitting that satisfies the preconditions of Theorem 4.1, we have an upper bound of $(k-1)2^{n-1}$. □

Note: When m is greater than $k-1$, the removal of an edge incident to the central vertex may not decrease the cubical dimension in both T_1 and T_2, so we cannot immediately use Theorem 4.1. Resolving these cases is still open.

4.4 k-stars

Next, we study $\mathrm{sat}(Q_n, S_{k_1\times k_2\times\cdots\times k_m})$. For lower bounds, we generalize the method used for a single star to get a rough lower bound. Then, for upper bounds, we use a variation on the hamming code to find surprisingly tight bounds. In particular, we first show, for 2-stars, how to improve the trivial upper bound given by Theorem 4.1. From there, we find improved upper bounds for all 3-stars and 4-stars, and then proceed to determine upper bounds for general k-stars with sufficiently large minimum degree.

We first provide a lower bound, best for when all degrees are large.

Theorem 4.13. *For $k_1, \ldots, k_m \ge 1$, $\mathrm{sat}(Q_n, S_{k_1\times\cdots\times k_m}) \ge (\min\{k_1,\ldots,k_m\} - 1 + o(1))\,2^{n-2}$.*

Proof. The proof is similar to that of Theorem 4.5 and is thus left to Appendix E. □

We now investigate upper bounds on the saturation numbers of these k-stars. For 2-stars $S_{k\times r}$, by Theorem 4.1, we trivially have an upper bound of $(\max\{k,r\}-1)2^{n-1}$. We show that, using a construction built around a hamming code in Q_{2^i+1}, we can improve this bound.

Before this, however, we present an important lemma that allows us to construct our saturated graphs.

Lemma 4.14. *For every k-regular bipartite graph H, there exists some subgraph G of H which is r-regular and bipartite for all nonnegative integers $r \leq k$.*

Proof. Note first that any subgraph G of a bipartite graph is necessarily bipartite, so we only need to find an r-regular subgraph. To do this, we invoke Hall's Theorem. By a simple application of this theorem, we can find a perfect matching within our k-regular subgraph. Removing all edges in this perfect matching, we are left with a $(k-1)$-regular bipartite subgraph. We can repeat this process $k - r$ times, and thereby end up with an r-regular, bipartite graph, as desired. □

Theorem 4.15. *For all positive integers k, r where $k > r$,* $\operatorname{sat}(Q_n, S_{k\times r}) \leq (r-1)2^n$.

Proof. Let κ be the smallest integer greater than k of the form $2^j - 1$. Now, consider a hamming code C on Q_κ, which is perfect by the definition of κ. From this, we will construct a $S_{\kappa\times r}$-saturated graph H on Q_κ. Begin by adding to H all incident edges to C. This creates $\frac{2^\kappa}{\kappa+1}$ vertices with degree κ. Now, note that the induced subgraph H' of H with vertex set $V(Q_n) \setminus C$ is 1-regular and bipartite, as it is a subgraph of Q_n. From here, our preconditions satisfied, we use Lemma 4.14 in reverse to add perfect matchings to H' until it is $(r-1)$-regular. Adding this to H, we have a subgraph in which $\frac{2^\kappa}{\kappa+1}$ vertices have degree κ and $\frac{\kappa\cdot 2^\kappa}{\kappa+1}$ vertices have degree $r-1$. From this, it is easy to see that H is $S_{\kappa\times r}$-saturated, as any non-edge must be incident to some vertex with degree $r-1$, thereby creating our 2-star, and it does not originally contain $S_{\kappa\times r}$, as there are no two adjacent vertices with degree κ.

To scale this up to Q_n, we simply need to consider the subgraph of $Q_\kappa \,\square\, Q_{n-\kappa}$ in which each vertex of $Q_{n-\kappa}$ contains H, and there are no edges between Q_κ's. In this subgraph of Q_n, all non-edges must be incident to some vertex with degree $r-1$, except those incident to two vertices in hamming codes. However, since any edge between degree κ vertices creates $S_{\kappa\times r}$, our subgraph remains saturated, giving us an upper bound of

$$2^{n-\kappa} \cdot \left(\kappa \cdot \frac{2^\kappa}{\kappa+1} + (r-2) \cdot \frac{\kappa \cdot 2^\kappa}{\kappa+1}\right) \leq (r-1) \cdot 2^n. \qquad \square$$

For convenience, we denote the construction in which we embed our saturated graph in a hypercube of dimension $\kappa = 2^i - 1 > k$ as a *κ-construction*. In particular, this allows us to both maintain a perfect hamming code and cover all but some set number of vertices in our k-star, since κ by definition is larger than $k_1, k_2, \ldots, k_m$.

With that aside, we move to the 3-star, 4-star, and special k-star cases. Because of space constraints, we leave the proofs of these results to Appendix E. Note that they contain some similar ideas to the construction of an upper bound for $S_{k\times r}$, although they require much more precision.

Theorem 4.16. *For all $k_1, k_2, k_3 \geq 2$, $\operatorname{sat}(Q_n, S_{k_1\times k_2\times k_3}) \leq (\min\{k_1, k_2, k_3\} - 1) \cdot 2^n$.*

Theorem 4.17. *For all $k_1, k_2, k_3, k_4 \geq 3$, $\operatorname{sat}(Q_n, S_{k_1\times k_2\times k_3\times k_4}) \leq (\min\{k_1, k_2, k_3, k_4\} - 1) \cdot 2^n$.*

Theorem 4.18. *Given $S_{k_1\times k_2\times\cdots\times k_m}$, let $\operatorname{emin}\{k_1, k_2, \ldots, k_m\} = (k_j, k_{j+1})$. Given that $m = 2^a + b$ for some integer $2 < b \leq 2^a$, then, if $b - 1 < j < 2^a$ and $\max\{k_j, k_{j+1}\} \geq \lfloor \log_2 m \rfloor$, $\operatorname{sat}(Q_n, S_{k_1\times k_2\times\cdots\times k_m}) \leq (\max\{k_j, k_{j+1}\} - 1) \cdot 2^n$.*

Theorem 4.19. *Consider $S_{k_1\times\cdots\times k_m}$ for $m = 2^a + 1$. If $\operatorname{emin}\{k_1, \ldots, k_m\} = (k_i, k_{i+1})$ for $i = 1$ or $i = m-1$, then, if $\max\{k_i, k_{i+1}\} \geq a$, $\operatorname{sat}(Q_n, S_{k_1\times\cdots\times k_m}) \leq (\max\{k_i, k_{i+1}\} - 1) \cdot 2^n$.*

Note that Theorem 4.18 works even if $\operatorname{emin}\{k_1, k_2, \ldots, k_m\} \neq (k_j, k_{j+1})$ as long as there do not exist two pairs (k_c, k_{c+1}) and (k_d, k_{d+1}), where $c \leq b - 1$ and $d \geq 2^a$, that satisfy $\max\{k_c, k_{c+1}\} \leq \max\{k_i, k_{i+1}\}$ and $\max\{k_d, k_{d+1}\} \leq \max\{k_i, k_{i+1}\}$. From this, it is easy to see that Theorem 4.18 gives a lower bound for most k-stars with length $m \neq 2^i + 1$, provided that the minimum degree is sufficiently large. Theorem 4.19 helps begin to resolve the case where $m = 2^i + 1$. Combined, they allow us to find upper bounds on the saturation number of almost every caterpillar.

Based on these results, we also suspect that, in fact, with a similar condition on the minimum degree, the following conjecture about all trees holds:

Conjecture 4.20. *Given a tree T with $\mathrm{emin}(T) = \delta$, $\mathrm{sat}(Q_n, T) < (\delta + C - 2) \cdot 2^n$ where C is the maximum distance from any given vertex to the longest central path in T.*

4.5 Beyond Caterpillars

In Section 4.4, it was conjectured that the arguments used for caterpillars could be extended to trees. In this section, we show that this indeed is the case for some trees, using the same type of argument for a class of trees that are not caterpillars.

In particular, we examine the *subdivided star* S_k^t, as described in [10], which consists of one central vertex with k legs, t of which contain two edges and the rest of which contain just one edge. We generalize this definition to the *r-subdivided star* $S_{k,r}^t$, where we instead have that t adjacent vertices to the central vertex have degree $r + 1$, while $k - t$ have degree r.

Theorem 4.21. *For all positive integers k, r, t satisfying $r < k$ and $t \leq \min\{k, \frac{\kappa-2}{2}\}$, if κ is some positive integer that satisfies $\kappa = 2^i - 1$ and $\kappa > k$, then $\mathrm{sat}(Q_n, S_{k,r}^t) \leq (r + t - 2) \cdot 2^n$.*

Proof. As usual, we consider a κ-construction and a hamming code C in Q_κ. In our saturated graph H, we add all edges incident to C and then use Lemma 4.14 to add perfect matchings until every other vertex has degree $r - 1$. From there, we pick $t - 1$ vertices adjacent to each $c \in C$ and add some incident edge between them. In particular, we pair hamming vertices so that there exist exactly $t - 1$ such vertices around each hamming vertex. At this point, let R be the set of vertices with degree r in H.

To make H saturated, we must also add edges between vertices in R, as otherwise these non-edges do not create our desired r-subdivided star. From each $v \in R$, there are at most

κ such additions. We greedily add these to maintain saturation. From here, we note that any other non-edge must be incident to some vertex with degree $r-1$, in which case $S^t_{k,r}$ is created. Therefore, our graph is saturated.

To scale this up to Q_n, we must ensure that our isomorphism of H does not match any u and v such that $u, v \in C \cup R$. However, because $|C \cup R|$ is at most half of the total number of vertices by construction, a symmetry argument allows us to create these isomorphisms. Therefore, our subgraph remains $S^t_{k,r}$-saturated. From here, all that remains to obtain an upper bound on the number of edges in our saturated graph,

$$2^{n-\kappa} \cdot (2^\kappa + (r-2) \cdot 2^\kappa + (t-1) \cdot 2^\kappa) = (r+t-2) \cdot 2^n. \qquad \square$$

This result implies that our general results for caterpillars might indeed generalize for all trees. In Appendix F, we derive another such bound for a more general class of trees.

5 Conclusion

In this paper, we examined the saturation number of many forbidden graphs in the hypercube. We first found general methods for finding lower and upper bounds on this saturation number of both general graphs and general trees. We continued by examining specific trees, and then used these to deduce upper bounds for the saturation number of sufficiently high-degree caterpillars. These bounds were surprisingly tight, and we conjecture that a similar argument can be extended to all trees.

In a larger context, our work provide expansive results on the saturation number of graphs in the hypercube, an area that is not particularly well studied. We hope that our bounds shed light on the methods that can be used to compute saturation numbers in the hypercube, and perhaps for any host graph. In the future, we hope to completely classify all trees from the work we have done here, and perhaps from there move to classifying cycles or subcubes. This problem, both in the hypercube and other host graphs, is fascinating and important in extremal graph theory and theoretical computer science, and we hope that our results are just the beginning in a complete characterization of the saturation number.

6 Acknowledgments

First and foremost, I would like to thank Chiheon Kim, my mentor, without whom this project would never have been possible. Furthermore, I would like to acknowledge Dr. Tanya Khovanova for enlightening conversations about my project and helpful guidance on the paper and presentation. I am also grateful to Dr. John Rickert for his invaluable guidance concerning my paper and presentation. In addition, I would like to extend my thanks to the Research Science Institute, the Center for Excellence in Education, and the MIT Math Department, all of whom made this amazing opportunity a reality for me. Finally, I would like to express my eternal gratitude to all of the sponsors who made my extraordinary experience here at RSI possible. Specifically, I would like to thank Dr. Frederick Y. Chen, Director of the Cardiac Surgery Laboratory at Brigham and Women's Hospital, Mr. John Yochelson, President of Building Engineering and Science Talent (BEST), Dr. Laura Stubbs, Director of Science and Technology Initiatives at the Department of Defense, Dr. Wolfgang Ketterle, John D. MacArthur Professor of Physics at MIT, Professor Hartley Rogers, Jr., Professor of Mathematics at MIT, Professor Philip A. Sharp, Chairman of the Department of Biology at MIT, Mr. Peter J. Barrer and Mrs. Judith A. Nichols, Mr. Peter L. Beebee, Mr. Piotr Mitros, Mr. Ray Prasad and Ms. Rima Pande, and Professor and Mrs. Neil Todreas.

References

[1] P. Turán. On an extremal problem in graph theory. *Matematikal és Fizikal Lapok*, 48: 436–452, 1941.

[2] P. Erdös. On some problems in graph theory, combinatorial analysis, and combinatorial number theory. *Graph Theory and Combinatorics*, pages 1–17, 1984.

[3] A. Bialostocki. Some ramsey type results regarding the graph of the n-cube. *Ars Combinatoria*, 16: 39–48, 1983.

[4] J. Balogh, P. Hu, B. Lidický, and H. Liu. Upper bounds on the size of 4- and 6-cycle-free subgraphs of the hypercube. *European Journal of Combinatorics*, 35: 75–85, 2014.

[5] F. Chung. Subgraphs of a hypercube containing no small even cycles. *Journal of Graph Theory*, 16: 273–286, 1992.

[6] Z. Füredi and L. Özkahya. On even-cycle free subgraphs of the hypercube. *Electronic Notes in Discrete Mathematics*, 34: 515–517, 2009.

[7] D. Conlon. An extremal theorem in the hypercube. *Electronic Journal of Combinatorics*, 17: X111, 2010.

[8] S. Choi and P. Guan. Minimum critical squarefree subgraphs of a hypercube. *Proceedings of the Thirty-Ninth Southeastern International Conference on Combinatorics, Graph Theory, and Computing*, 189: 57–64, 2008.

[9] J. R. Johnson and T. Pinto. Saturated subgraphs of the hypercube. 2014. Preprint, `http://arxiv.org/abs/1406.1766`.

[10] J. Faudree, R. Faudree, R. Gould, and M. Jacobson. Saturation numbers for trees. *The Electronic Journal of Combinatorics*, 16: R91, 2009.

[11] Q. Stout. Packing in hypercubes. Presented at 21st Southeastern International Conference on Combinatorics, Graph Theory, and Computing, Boca Raton, FL, 1990, `http://www.eecs.umich.edu/~qstout/abs/hyppack.html`.

[12] P. Weichsel. Dominating sets in n-cubes. *Journal of Graph Theory*, 18: 479–488, 1994.

Appendix A: General Results

In this appendix, we prove our lower bound on the saturation number for general graphs G and also present a lemma that shows an alternative method of getting an upper bound of $O(2^n)$ for the saturation number of any graph G.

Proof of Theorem 3.1. Denote our G-saturated graph by H. We first consider the expression

$$\sum_{uv\in E(Q_n)} \deg(u)+\deg(v). \tag{5}$$

We can evaluate (5) in two ways. First, we can evaluate it directly. We know that each of u and v have n edges incident to them in Q_n, so in fact (5) is equivalent to

$$2n\sum_{u\in V(Q_n)} \deg(u) = 4ne(H),$$

as $\sum_{u\in V(Q_n)} \deg(u)$ is just $2e(H)$.

However, we can also evaluate (5) in another way. In particular,

$$\sum_{uv\in E(Q_n)} \deg(u)+\deg(v) = \sum_{uv\in E(H)} \deg(u)+\deg(v) \;+ \sum_{uv\in E(Q_n)\setminus E(H)} \deg(u)+\deg(v). \tag{6}$$

Here, we first evaluate

$$\sum_{uv\in E(H)} \deg(u)+\deg(v). \tag{7}$$

We can split (7) up, and see that, because we are in H, the number of incident edges to u is $\deg(u)$. Thus, (7) is equivalent to $2\displaystyle\sum_{u\in V(H)} \deg^2(u)$. From here, using the Cauchy-Schwartz inequality, we can bound (7) from below, getting that

$$2\sum_{u\in V(H)} \deg^2(u) \geq 2\cdot\frac{(2e(H))^2}{2^n} = \frac{e(H)^2}{2^{n-3}}. \tag{8}$$

For the second part of (6), $\sum_{uv\in E(Q_n)\setminus E(H)} \deg(u)+\deg(v)$, we must use the fact that $\mathrm{emin}(G)=\delta+1$. We are considering edges that are not already in H, and therefore need to add a copy of G to H. Since this requires the portion of H to which we are adding the edge to have an isomorphism of $G-e$ for some edge $e\in G$, we know that at least one of the vertices u and v must have degree δ. Therefore, we can get the rough bound $\deg(u)+\deg(v)\geq\delta$. From this, we get that

$$\sum_{uv\in E(Q_n)\setminus E(H)} \deg(u)+\deg(v) \geq \delta(n2^{n-1}-e(H)). \tag{9}$$

Combining (8) and (9) in (6), we get

$$4ne(H) \geq \frac{e(H)^2}{2^{n-3}} + \delta(n2^{n-1}-e(H)).$$

Multiplying by 2^{n-3} and collecting terms, this is equivalent to

$$e(H)^2 - (4n+\delta)2^{n-3}e(H) + \delta n 2^{2n-4} \leq 0.$$

Factoring,

$$(e(H)-\delta\cdot 2^{n-3})(e(H)-n2^{n-1}) \leq 0.$$

Since $\delta\cdot 2^{n-3} < n2^{n-1}$, this implies that $e(H)\geq\delta\cdot 2^{n-3}$, so we are done. $\square$

Note: As a corollary to Theorem 3.1, we can show that the saturation number of all connected graphs G with $e(G)>1$ has a lower bound that is $O(2^n)$, $\mathrm{emin}(G)$ cannot be 0.

Our next result, which we solely include in the appendix, allows us to construct an upper bound on the saturation number of almost any graph in the hypercube. To do this, we use the fact that we can create a saturated subgraph in Q_{i+1} by taking two copies of a saturated subgraph in Q_i, and then adding some number of edges between them. We show that if the number of edges added is consistently small enough, then we have an upper bound on the order of 2^n.

Lemma A.1. *Consider a subgraph G of Q_k, and let $A_k, A_{k+1}, \ldots, A_n$ be recursively constructed subgraphs in the hypercube, where A_i for $i > k$ is formed by taking two copies of A_{i-1} in $(0, \{0,1\}^{i-1})$ and $(1, \{0,1\}^{i-1})$ and adding some number a_i of edges between them. If, for some sufficiently large j, $\frac{a_m}{a_{m-1}} < 2$ for all $m \geq j$, then $\mathrm{sat}(Q_n, G) < c \cdot 2^n$ for some constant c.*

Proof. First, consider A_j for our sufficiently large integer j. It contains some constant number C of edges. Now, we know that the sequence $a_j, a_{j+1}, \ldots, a_n$ has the property that the ratio between consecutive terms is less than 2. In particular, let this ratio be bounded above by 2^ϵ, where $0 < \epsilon < 1$. Then the sequence $a_j, 2^\epsilon a_j, 2^{2\epsilon} a_j, \ldots, 2^{(n-j)\epsilon} a_j$ has the property that each of its terms is greater than the corresponding term in our original sequence. Now, note that our sequence of saturated subgraphs satisfies the recursion $|A_m| = 2|A_{m-1}| + a_m$ for $m > j$. We can substitute $2^{(m-j)\epsilon}$ for a_m in this sequence to get an upper bound. This gives us

$$|A_n| \leq 2^{(n-j)\epsilon} + 2|A_{n-1}| \leq 2 \cdot 2^{(n-j)\epsilon} + 4|A_{n-2}| \leq \cdots \leq (n-j)2^{(n-j)\epsilon} + C2^{n-j} < n2^{n\epsilon} + C2^n.$$

Now, consider the limit

$$\lim_{n\to\infty} \frac{n2^{n\cdot\epsilon}}{2^n} = \frac{n}{2^{(1-\epsilon)\cdot n}}.$$

Using l'Hôpital's rule, this limit is equivalent to

$$\lim_{n\to\infty} \frac{1}{\ln 2 \cdot (1-\epsilon) \cdot 2^{(1-\epsilon)\cdot n}} = 0$$

as the denominator still increases with n, while the numerator is constant. This tells us that

$n2^{n\epsilon} = o(2^n)$, so we have that

$$n2^{n\epsilon} + C2^n < (C+1)2^n.$$

Thus, since we have an upper bound that is $O(2^n)$, we are done. $\square$

NOTE: This proof does not cover the number of edges being 0 because of the infinities it creates in the limit. However, it is clear that if the number of edges we need to add is 0, the bound is trivially $O(2^n)$, as we are simply doubling the number of edges. Similarly, our proof does not cover $\epsilon = 0$, but this is trivially also on the order of 2^n since $2^{n\epsilon} = 1$.

Appendix B: Paths

In this appendix, we present a proof of Theorem 4.4, which gives an upper bound on the saturation number for paths of length k.

Proof of Theorem 4.4. We begin by describing constructions for the even and odd cases, enumerate the number of edges in each, and then show why each of them is P_k-saturated. Before we begin, however, it is necessary to define a sequence of vertices $w_1, w_2, w_3, \ldots, w_{2^{k-1}}$ on Q_k, where w_1 is any vertex, and w_i is chosen randomly from the set of remaining vertices S that satisfy, for all $s \in S$, $|w(s) - w(w_1)| \equiv 0 \pmod 2$.

We now split the proof into two separate cases.

Case 1: $k \equiv 1 \pmod 2$

All odd integers can be represented in the form $k = 2^i + j$, where $j = 2m + 1$ and $0 < j < 2^i$. Now, consider a subgraph H of Q_{i+1} that contains all of the edges in $(0, \{0,1\}^i)$, all of the edges between $(0, \{0,1\}^i)$ and $(1, \{0,1\}^i)$ that are not incident to vertices in

$w_1, w_2, \ldots, w_{2^{i-1}-m}$, and all of the edges in $(1, \{0,1\}^i)$ after the deletion of all edges containing $w_1, w_2, \ldots, w_{2^{i-1}-m}$.

The number of edges in such a construction, scaled up to Q_n, is, by simple computation,

$$2^{n-i-1} \cdot \left((i+1)2^i - (i+1)(2^{i-1} - m)\right),$$

which simplifies to

$$(i+1)2^{n-2} + (i+1)m \cdot 2^{n-i-1},$$

our desired upper bound.

Thus, we just need to show that H is P_k-saturated. To do this, first notice that, by Lemma 4.3, H must be P_k free, since $2^{i+1} - 2(2^{i-1} - m) = 2^i + 2m = 2^i + j - 1$ is the maximum length of such a path. To show that the addition of any non-edge creates P_k, we must show that exactly 2^i vertices in our construction are endpoints. To do this, note that, by a simple parity argument, a path of length 2^{i-1} in $(0, \{0,1\}^i)$ ending at an vertex of even weight in must begin at a vertex of odd weight. Further note that, once this path has been extended to length 2^i into $(1, \{0,1\}^i)$, the path, beginning at a vertex of odd weight, must also end at a vertex of odd weight, since the path in question has even length. And, by the symmetry of our edge removal, these vertices can be chosen from any among the set of vertices in $(1, \{0,1\}^i$. This leads to a grand total of $2^{i-1} + 2^{i-1} = 2^i$ endpoints, which is sufficient by Lemma 3.3 to imply our result.

Case 2: $k \equiv 0 \pmod 2$

All even integers can be represented in the form $k = 2^i + j + 1$, where $j = 2m + 1$ and $0 < j < 2^i$. For $j = 2^i - 1$, our construction simply consists of the entire graph Q_{i+1}. Otherwise, it consists of both the subgraph H of Q_{i+1} that is P_{k-1}-saturated and all edges between $w_1, w_2, \ldots, w_{2^{i-1}-m}$ and corresponding vertices on $(0, \{0,1\}^i)$.

The number of edges in such a construction is just $2^{n-i-1} \cdot (2^{i-1} - m)$, the number of edges incident to some w_i in $(1, \{0,1\}^i)$, added with the number of edges in our P_{k-1}-saturated graph. This gives an upper bound of:

$$(i+2) \cdot 2^{n-2} + im \cdot 2^{n-i-1},$$

as desired.

Thus, we just need to show that H is P_k-saturated. To do this, we first show that H is P_k-free. To show this, note that we are simply adding edges between vertices of odd weight in $(0, \{0,1\}^i)$ and vertices of even weight in $(1, \{0,1\}^i)$. This adds at most two edges to the longest path in the previous construction. We wish to show that, even when both of these edges are added, the path's maximal length increases by at most 1. To do this, first note that one of these edges must begin the path, and one must end the path. Removing these two edges for a moment, this means that the beginning and final vertex in the resultant path of $k-2$ edges must be of the same parity on $(0, \{0,1\}^i)$. However, as seen in Lemma 4.3, when the number of edges between $(0, \{0,1\}^i)$ and $(1, \{0,1\}^i)$ is even, the maximal length of a path is actually one less than the overall maximum, so the overall maximal path still has length $k-3+1 = k-1$. Thus, our graph is P_k-free. From here, notice that any non-edge in H incident to one of these vertices of even weight necessarily creates P_k by appending an edge to the aforementioned P_{k-1}. This covers 2^{i-1} endpoints: we need 2^{i-1} more. For these, note that vertices of odd weight in $(1, \{0,1\}^i)$ were also endpoints in the odd construction with P_{k-2}'s ending at some vertex of odd weight in $(0, \{0,1\}^i)$. Since we are now appending an edge to this, these vertices of odd weight in $(1, \{0,1\}^i)$ are still endpoints. Therefore, we have 2^i endpoints, sufficient to show our upper bound by Lemma 3.3. Thus, we are done. $\square$

Appendix C: Stars

In this appendix, we show how to improve the upper bound on the saturation number of stars given by Theorem 4.1 by eliminating a large portion of the edges in this construction.

Proof of Theorem 4.6. It is trivial to use Theorem 4.1 to get a bound of $(k-1)2^{n-1}$, as the removal of any given edge from a star with k edges allows it to be placed in Q_{k-1}: in other words, the cubical dimension is lowered. However, we want to improve this bound. In particular, let us first describe the construction that achieves this upper bound, and then improve it.

To get the original bound, consider the star with k edges. The saturated subgraph H that achieves the trivial upper bound in Q_n is the one that only takes edges in the Q_{k-1}'s on the vertices of $Q_{k-1} \square Q_{n-k+1}$. This yields $(k-1) \cdot 2^{k-2} \cdot 2^{n-k+1} = (k-1)2^{n-1}$ edges.

However, there is a better way. Consider, within H, a vertex v in the $(k-1)$-dimensional hypercube $(0, \{0,1\}^{k-1})$ and its counterpart v' on $(1, \{0,1\}^{k-1})$, where all of the remaining $n-k$ vertices are, for the moment, fixed. Currently, we have no edges between such vertices; just edges within the two Q_{k-1}'s. However, now consider two adjacent vertices to v on $(0, \{0,1\}^{k-1})$, x and w, and two to v' in the same directions on $(1, \{0,1\}^{k-1})$, x' and w'. If we delete the edges vx and $v'w'$, and add the edge vv', the overall Q_k remains saturated. To see this, note that vx still creates S_k because v has degree $k-1$, xx' creates S_k because x' has degree $k-1$, and ww' creates S_k because w has degree $k-1$. Therefore, our S_k-saturation is maintained, but we have one fewer edge. We can continue this removal for other vertices while noting that both our pair and adjacent vertices in later deletions cannot contain v, v', w, w', x, or x'. This implies that our upper bound is simply the densest possible vertex-disjoint packing of Q_{k-1} with P_2. Stout [11] showed that such a vertex disjoint packing goes to density 1 as k increases, so the number of such paths tends toward $\frac{2^k}{3}$. Thus, since each deletion removes exactly one edge, we can delete $\frac{2^k}{3}$ edges from our original construction, giving a bound of $(k-1)\cdot 2^{n-1} - \frac{2^k}{3}\cdot 2^{n-k} = \big(k - \frac{5}{3} + o_k(1)\big)\cdot 2^{n-1}$. Therefore, we are done. $\square$

Appendix D: Generalized Stars

In this appendix, we show how to prove the specific upper bounds on the saturation numbers of $GS_{k,2}$, $GS_{k,3}$, and $GS_{k,5}$ which rely on particular constructions within Q_k.

Proof of Theorem 4.7. Consider the subgraph H of Q_k consisting of all edges in $(0, \{0,1\}^{k-1})$ and all edges between $(0, \{0,1\}^{k-1})$ and $(1, \{0,1\}^{k-1})$. Note that, for H to contain $GS_{k,2}$, there must exist some vertex with degree k in the former $(0, \{0,1\}^{k-1})$ with a path of length 2 emanating from all of its vertices. However, of course, given a vertex v in $(0, \{0,1\}^{k-1})$, there is no such path in the first direction, in which there is only a path of length 1. However, there is such a P_2 in every other direction. For example, in the kth direction, it is only necessary to traverse the edge in the kth direction from v in $(0, \{0,1\}^{k-1})$, and then traverse the first direction into $(1, \{0,1\}^{k-1})$. H is also $GS_{k,2}$-saturated. We see this in the fact that any non-edge added must be in $(1, \{0,1\}^{k-1})$. Any such edge, thus, creates a path of length 2, since it must be incident to one of the original edges between $(0, \{0,1\}^{k-1})$ and $(1, \{0,1\}^{k-1})$. Since all other directions already have a path of length 2, this creates $GS_{k,2}$. From here, all that is necessary is to place H at all of the vertices of $Q_k \,\square\, Q_{n-k}$, where saturation is clearly maintained, which yields as an upper bound

$$2^{n-k} \cdot \big((k-1)2^{k-2} + 2^{k-1}\big) = 2^{n-k} \cdot (k+1) \cdot 2^{k-2} = (k+1) \cdot 2^{n-2}.$$

$\square$

Proof of Theorem 4.8. Consider the subgraph H of Q_k consisting of all edges in $(0, \{0,1\}^{k-1})$, all edges between $(0, \{0,1\}^{k-1})$ and $(1, \{0,1\}^{k-1})$, and all edges in the kth direction in $(1, \{0,1\}^{k-1})$.

Note that for H to contain $GS_{k,3}$, there must be some vertex with degree k in $(0, \{0,1\}^{k-1})$ with a path of length 3 in all directions. In particular, this vertex must have a path of length 3 emanating in the first direction. Once this edge travels to the hypercube $(1, \{0,1\}^{k-1})$, however, it can only travel in the kth direction, since these are the only edges on that hypercube. From there, it cannot travel again in the kth direction, so it can only traverse the first direction back to $(0, \{0,1\}^{k-1})$. But this endpoint of our P_3 is adjacent to the original vertex in the kth direction, and therefore our original vertex cannot have a path of length 3 in the kth direction without crossing this vertex twice, which is not allowed.

From this, it is clear that there is no $GS_{k,3}$ in H. However, also notice that otherwise, $k-1$ paths of length 3 can be created in all but the first direction by travelling in each direction, traversing the first direction into $(1, \{0,1\}^{k-1})$ and then travelling along the kth direction. The only exception is the path that starts in the kth direction, which we keep on the original hypercube. This ensures that whenever any edge is added in $(1, \{0,1\}^{k-1})$, some path of length 3 in the first direction is completed, and we have $GS_{k,3}$.

Thus, our subgraph is $GS_{k,3}$-saturated. All that remains is to enumerate the number of edges when this is scaled up to Q_n, where saturation is clearly maintained, which yields as an upper bound

$$2^{n-k} \cdot \left((k-1)2^{k-2} + 2^{k-1} + 2^{k-2}\right) = 2^{n-k} \cdot (k+2) \cdot 2^{k-2} = (k+2) \cdot 2^{k-2}.$$

$\square$

Proof of Theorem 4.9. Consider the subgraph H of Q_k consisting of all edges in $(0, \{0,1\}^{k-1})$, disjoint Q_2's covering all vertices in $(1, \{0,1\}^{k-1})$, and edges between $(0, \{0,1\}^{k-1})$ and $(1, \{0,1\}^{k-1})$ consisting of exactly two vertices in each disjoint Q_2.

We claim that H is saturated. To show this, we first show that it must necessarily be free of H. To do this, note first that, in the last $k-1$ directions, a path of length 5 can evidently be found in H in all directions. However, in the first direction, consider any path. The maximal path, evidently, consists first of traversing one of the edges in the first direction into $(1, \{0,1\}^{k-1})$. Then, we must necessarily return to some other vertex that has any edge in the first direction, for which the only candidate is an adjacent vertex to where we are now by construction. Traversing a path of length 3 in the square, we arrive at this vertex. However, from here, the only way to achieve a path of length 5 is to traverse the edge in the first direction back into $(1, \{0,1\}^{k-1})$. However, this is necessarily adjacent to our original vertex in some direction, and therefore cuts off one of the other paths. This means that indeed there cannot be any $GS_{k,5}$ in H.

Now, the addition of any edge in Q_k to H creates $GS_{k,5}$. To do this, first note that any edge added in $(1, \{0,1\}^{k-1})$ between squares necessarily creates some P_5 from a vertex in $(0, \{0,1\}^{k-1})$ in the first direction, and therefore creates $GS_{k,5}$. In particular, this path consists of at least the edge in the first direction from this vertex in $(0, \{0,1\}^{k-1})$, some nonnegative path within the current square in $(1, \{0,1\}^{k-1})$, the edge between this square and another square, and then a maximal path of length 3 in that square. This has length at least $1+1+3=5$, and therefore we are done.

The only other possibility is an edge added between vertices in $(0, \{0,1\}^{k-1})$ and $(1, \{0,1\}^{k-1})$. In this case, if an edge of this type is added, then there must exist some Q_2 in $(1, \{0,1\}^{k-1})$ where opposite vertices have an edge in the first direction. Let these vertices be v and w, and let their corresponding vertices in $(0, \{0,1\}^{k-1})$ be v' and w'. Now, consider the path $v' - v - a - w - w' - b$, where a is the vertex between v and w in the Q_2 and b is some adjacent vertex to w' in $(0, \{0,1\}^{k-1})$ that is not adjacent to v. This is a path of length 5, and note from here that we can still find $k-1$ paths of length 5 from v by traversing the direction and then travelling into $(1, \{0,1\}^{k-1})$ to complete the path of length 5. The only possible exceptions exist in the directions of the Q_2 in question. However, in these cases, it is clear that we can simply construct paths of length 5 in $(0, \{0,1\}^{k-1})$ while avoiding w' and

b. Therefore, $GS_{k,5}$ is created and our graph is saturated.

To scale this up to Q_n, consider the subgraph of Q_n consisting of placing copies of our saturated graph at the vertices of $Q_k \square Q_{n-k}$, with no initial edges between Q_k's. Any edge within a Q_k evidently creates a copy of $GS_{k,5}$ because these subgraphs are saturated, and the overall graph itself is G-free as no initial edges are added. Now, consider the final case, when an edge is added between Q_k's. In this case, it either connects two full Q_{k-1}'s or connects two Q_{k-1}'s consisting of disjoint Q_2's. In the former case, we simply take $k-1$ paths of length 5 starting at some vertex in one of the Q_{k-1}'s, and then add a path of length 4 after traversing the added edge into the other Q_{k-1} to create $GS_{k,5}$. The more complicated case is the latter. Let u and v be the incident vertices to this edge, and let S and S' be the respective disjoint Q_2's that they lie in. Finally, let b be one of the vertices in S that has an edge in the first direction, and let b' be the other endpoint of its edge in one of the full Q_{k-1}'s. From this information, we can construct our $GS_{k,5}$ by taking the path $b' - b - u - v - a$, where a is the endpoint of a path of length 3 in S' starting at v. Note that this path has minimum length 5 when $b = u$, so $GS_{k,5}$ is created. Thus, our constructed subgraph of Q_n is $GS_{k,5}$-saturated.

From here, all that remains is to enumerate the number of edges in our subgraph to get an upper bound, which gives, as desired,

$$2^{n-k} \cdot \big((k-1)2^{k-2} + 2^{k-2} + 2 \cdot 2^{k-2}\big) = (k+2) \cdot 2^{n-2}.$$

□

From here, we present a general result on the saturation number of generalized stars which, because of space constraints, could not be stated in our paper.

Theorem D.1. *For $m = 2^i + 1$ and $k > m$,* $\operatorname{sat}(GS_{k,m}) \leq (k+1+i) \cdot 2^{n-2}$

Proof. Consider the subgraph H of Q_k consisting of all edges in $(0, \{0,1\}^{k-1})$, and disjoint

Q_i's covering all vertices in $(1, \{0,1\}^{k-1})$.

We first show that H is free of $GS_{k,m}$. Note that, in the last $k-1$ directions, a path of length m can always be found even in $(0, \{0,1\}^{k-1})$. However, in the first direction, there are no edges, and therefore there can be no path of length m. Thus, it does not contain $GS_{k,m}$.

From here, we greedily add edges in the first direction to reach a saturated state, possible by Lemma 4 in [9]. This adds at most 2^{k-1} edges. One important detail to note here is that at least one edge must be added incident to every disjoint Q_i in $(1, \{0,1\}^{k-1})$, as otherwise this edge can be added to H later and not create $GS_{k,m}$, as $1 + 2^i - 1 = 2^i < m$. At this point in time, our graph must necessarily be saturated in Q_k. To scale this up to Q_n, consider placing our graph H at all vertices of $Q_k \,\square\, Q_{n-k}$, with no initial edges between Q_k's. Any edge within H, because it is saturated, creates $GS_{k,m}$, and H itself is $GS_{k,m}$-free. Furthermore, because there are no edges between Q_k's, the graph itself must necessarily be free of $GS_{k,m}$. The remaining case is whether any edge between Q_k's adds $GS_{k,m}$. If this edge is between full Q_{k-1}'s in H, then we simply pick $k-1$ paths of length m in one of the hypercubes, and then traverse this edge and four others in the other hypercube to create $GS_{k,m}$. The second case, in which the edge is between Q_{k-1}'s with only disjoint Q_i's, is a little more complicated. Let B and B' represent the Q_i's that this edge is added incident to. Then there exists some vertex $v \in B$ that contains an edge in the first direction, say connected to u in a full Q_{k-1}. Furthermore, let the endpoints of our edge be $b \in B$ and $b' \in B'$. To construct $GS_{k,m}$, first take $k-1$ paths of length m starting at u in the full Q_{k-1}. Then, traverse $u - v$ into our non-full Q_{k-1}, travel from v to b, travel from b to b', and then traverse a path of length $2^i - 1$ in B' starting at B. The minimal case here is when $v = b$, in which case the path is of length $1 + 1 + 2^i - 1 = 2^i + 1 = m$, which is sufficient. Therefore, H is saturated.

From here, all that remains is to enumerate the total number of edges in H. This gives an upper bound of

$$2^{n-k} \cdot \left((k-1)2^{k-2} + 2^{k-1} + i \cdot 2^{k-2}\right) = (k+1+i) \cdot 2^{n-2}.$$

$\square$

Appendix E: Caterpillars

In this appendix, we presents proofs of our general upper bounds on the saturation number for many caterpillars. Notably, all of these proofs rely on the existence of some dominating set of Q_n to facilitate their construction.

Proof of Theorem 4.13. First, without loss of generality, assume that $k_1 = \min\{k_1, k_2, \ldots, k_m\}$. Now, consider any vertex incident to a non-edge in our saturated graph. If this non-edge is added, it must complete some k-star. Therefore, at minimum, if the vertex itself does not have degree at least k_1, it must be incident to a vertex with degree at least $k_1 - 1$. With this aside, consider some subgraph H of Q_n which is $S_{k_1 \times k_2 \times \cdots \times k_m}$-saturated, and define V_i as the number of vertices with $\deg(v) = i$, for $i < k_1 - 1$. Note first that any vertex x with degree i must have at least $n - i$ neighbors not in H with degree at least $k_1 - 1$. From here, consider the number of vertices degree $k_1 - 1$ counted in such a search,

$$\sum_{i=0}^{k_1-2} (n-i)V_i.$$

Since every vertex with degree $k_1 - 1$ is counted at most $n - k_1 - 1$ times in this summation, we have that

$$\sum_{i=0}^{k_1-2} (n-i)V_i \leq (n - k_1 + 1)V_{k_1-1}. \tag{10}$$

At this point, it is also useful to note that

$$\sum_{i=0}^{k_1-1} iV_i \leq 2e(H),$$

as each edge in H is counted twice in the summation over all k. This leads us to the following expression as a manipulation of (10):

$$\sum_{i=0}^{k_1-2} (n-i)V_i \leq n \sum_{i=0}^{k_1-2} V_i - \sum_{i=0}^{k_1-2} iV_i. \tag{11}$$

From here, we can use our already determined bounds to simplify (11) to

$$n(2^n - V_{k_1-1}) - (2e(H) - (k_1-1)V_{k-1}) \leq (n - k_1 + 1)V_{k_1-1}.$$

Grouping terms, we have that

$$n2^n \leq 2e(H) + 2(n-k+1)V_{k_1-1}. \tag{12}$$

From here, we need some relationship between $e(H)$ and V_{k_1-1}. But, in this case, we know that

$$2e(H) \geq \sum_{i=0}^{k_1-1} iV_i \geq (k_1-1)V_{k_1-1},$$

so $V_{k_1-1} \leq \frac{2e(H)}{k_1-1}$. This in turn simplifies (12) to

$$n2^n \leq \frac{4(n-k_1+1)}{k_1-1} e(H),$$

or

$$e(H) \geq \frac{k_1-1}{2n-k_1+1} n2^{n-1} \sim (k_1 - 1 + o(1))2^{n-2}.$$

Thus, we are done. $\square$

Proof of Theorem 4.16. Let us split this proof into three distinct cases, assuming without loss of generality that $k_1 \geq k_3$. First, we consider the case $k_1 \geq k_2 \geq k_3$.

Within this, we consider both $k_2 = k_3$ and $k_2 > k_3$. For $k_2 = k_3$, we can otherwise represent our star by $S_{k_1 \times k_3 \times k_3}$. We create the exact same κ-construction as in the 2-star case. Note that it is indeed clear that, with the addition of any non-edge, not only in this case do we create $S_{\kappa \times k_3}$, but we also create $S_{\kappa \times k_3 \times k_3}$, since the non-edge added must connect vertices with degree at $k_3 - 1$. The only catch in this case is that, when scaling up to Q_n, we must actually take a translated version of H at odd weight vertices to ensure $S_{k_1 \times k_3 \times k_3}$ saturation, as no two vertices in a hamming code can be adjacent to one another. However, this process complete, our upper bound on the saturation number remains exactly the same, $(k_3 - 1) \cdot 2^n$.

When $k_2 > k_3$, we actually construct, after scaling up with our κ-construction, an $S_{\kappa \times \kappa \times k_3}$-saturated graph. The reason this implies $S_{k_1 \times k_2 \times k_3}$ saturation is that no three vertices with degree κ are connected, so there exists no 3-star originally. Furthermore, since $k_2 < \kappa$, creating a $S_{\kappa \times \kappa \times k_3}$ implies the existence of the smaller star.

With that aside, note that Weichsel [12] showed that, for $n = 2^i - 1 + k$, there exists a perfect dominating set S of Q_n where S consists of disjoint Q_j. We use here the case $j = 1$, changing our κ-construction accordingly. Therefore, we now have a dominating set of Q_1's, say C. Now, we construct our saturated subgraph H. First, we add all edges incident to C. Then, we use Lemma 4.14 again to find an $(k_3 - 2)$-regular bipartite graph among the remaining vertices. This leads to a subgraph where all vertices in C have degree κ, and all other vertices have degree $k_3 - 1$. We now must show that H is saturated in Q_κ. First, note that any non-edge must be incident to some vertex v with degree $k_3 - 1$, since all edges within the dominating sets are already in H. Once this non-edge is added, v has degree r. Furthermore, we can guarantee, since C dominates, that v is adjacent to exactly one element in C. Let this edge be $e = xy$. After the non-edge is added, $v - x - y$ consists of connected nodes with degree k_3, κ, and κ, respectively. Since this is exactly $S_{\kappa\times\kappa\times k_3}$, it remains to show that H is $S_{\kappa\times\kappa\times k_3}$-free. But note here that, for H to contain this 3-star, three vertices with degree greater than $k_3 - 1$ would have to be adjacent. However, the only such vertices satisfying this degree constraint are in C, and these only occur in pairs of exactly 2 vertices. Therefore, H is free of $S_{\kappa\times\kappa\times k_3}$. The only step that remains is to construct up to Q_n. In this case, copying onto the vertices of $Q_\kappa \,\square\, Q_{n-\kappa}$ almost works, except we again need to shift the vertices of our dominating set. In particular, we translate our Q_1 dominating set in two separate directions, creating a new dominating set C_0. This clearly cannot map any vertices to their neighbors, and, since all dominating sets have minimum distance 3, ensures that C_0 and C are disjoint and nonadjacent on separate hypercubes. Thus, by placing these two alternating isomorphisms of H at odd and even weight vertices of $Q_{n-\kappa}$, we have a construction that is saturated. All that remains now is to enumerate the number of edges in our construction. Using that the number of elements of C, from [12], is $\frac{2^{\kappa-j}}{\kappa-j+1} = \frac{2^{\kappa-1}}{\kappa}$, the total number of edges in such a construction, and therefore our upper bound, is

$$2^{n-\kappa} \cdot \left((2\kappa - 1) \cdot \frac{2^{\kappa-1}}{\kappa} + (k_3 - 2) \cdot \frac{(\kappa - 1)2^\kappa}{\kappa} \right) \leq (k_3 - 1)2^n.$$

Note quickly that, in this case, we could also construct a saturated graph by copying H and then connecting each vertex of $Q_{n-\kappa}$ to exactly one neighbor. Since, in this case, hamming code vertices would be connected exactly once, adding another edge would create $S_{\kappa\times\kappa\times k_3}$. However, since this increases the number of edges, we decided to use the former construction.

With $k_1 \geq k_2 \geq k_3$ aside, we consider $k_1 > k_3 > k_2$. Here, we again use a κ-construction, constructing a $S_{\kappa\times k_2\times\kappa}$-saturated graph, which implies $S_{k_1\times k_2\times k_3}$-saturation provided that the hamming codes we use are not adjacent in more than one vertex.

With that aside, we consider two hamming codes on Q_n: C and $D = C + v_1$, where v_1 is the basis vector in the 1st direction. Notably, these two hamming codes share no vertices, and every vertex in $V(Q_n)\backslash(C \cup D)$ is adjacent to exactly one vertex in both C and D. From here, we construct our saturated graph H. As usual, we add all edges incident to C and D. Then, we use Lemma 4.14 on the remaining vertices, possible by symmetry, to find an $(k_2 - 3)$-regular graph among the remaining vertices. This leads to an overall subgraph H where all edges in C or D have degree κ and the rest have degree $k_2 - 1$.

Next, it is necessary to show saturation. For this, first note that the graph is clearly $S_{\kappa \times k_2 \times \kappa}$-free, since otherwise there would need to be three adjacent vertices of degree κ, clearly not possible since we only consider two adjacent hamming codes. Furthermore, note that any edge in Q_κ added to this subgraph necessarily is incident to some vertex with degree $k_2 - 1$. Let this vertex be v. Then v has some incident vertex in C, say c, and some incident vertex in D, say d. Taking the path $c - v - d$ and all incident edges, we almost have $S_{\kappa \times k_2 \times \kappa}$. However, note that because c and d are exactly distance 2 away, the stars around them must share some vertex. However, because our κ-construction is made such that $\kappa > k_1$, our graph is still $S_{k_1 \times k_2 \times k_3}$-saturated, if not $S_{\kappa \times k_2 \times \kappa}$-saturated.

Therefore, the one step that remains is to construct this up to Q_n. In this case, as usual, copying onto the vertices of $Q_\kappa \square Q_{n-\kappa}$ makes the graph almost saturated. To fix the problem that arises when hamming codes vertices are adjacent, we simply shift C and D in the 2nd direction, deriving two completely disjoint hamming codes from C and D. We place this translation of H at even weight vertices and H at odd weight vertices, and through this maintain saturation. Thus, to find an upper bound on the saturation number, all that remains to enumerate the number of edges in our construction, which is exactly:

$$(2\kappa - 1) \cdot \frac{2^\kappa}{\kappa + 1} + (k_2 - 3) \cdot \frac{\kappa 2^\kappa - 2^\kappa}{\kappa + 1} \leq (k_2 - 1) \cdot 2^n.$$

Our final case is $k_2 > k_1 \geq k_3$. However, it is clear that we can just use the same construction as in the $S_{k_1 \times k_1 \times k_3}$ case (where, in this case, k_2 functions as k_1) and achieve the same bound of $(k_3 - 1) \cdot 2^n$. Thus, having completed all cases, we are done. $\square$

Proof of Theorem 4.17. First, consider the case where $k_1 = \min\{k_1, k_2, k_3, k_4\}$ or $k_4 = \min\{k_1, k_2, k_3, k_4\}$. Both of these minimal cases evidently reduce to the $S_{k \times k \times r \times k}$ case, given that we are constructing hamming codes with degree $\kappa > k$. Therefore, start by considering the smallest $\kappa > k$ which satisfies $\kappa = 2^j - 1$ for some positive integer j. We will attempt to find a $S_{\kappa-1 \times \kappa-1 \times r \times \kappa-1}$-saturated graph, which in turn implies our result. Denote this saturated graph by H. Begin by considering two hamming codes on Q_n, C and $D = C + v_1$, or a translated copy of C. Now, add to H all edges incident to C and D, thereby giving all of them degree κ. Then, as usual, use Lemma 4.14 to find and add $r - 3$ perfect matchings, thereby creating an $(r-1)$-regular graph among the vertices not in $C \cup D$. This finishes the construction of H.

Now, we show that H is saturated. It is evidently free of $S_{k \times k \times r \times k}$, as there are no four vertices of degree k adjacent to one another from our construction. Now, consider some non-edge in Q_n. If added, it must be incident to some vertex v, which has degree $r-1$. This edge added, consider two neighbors of v, $c \in C$ and $d \in D$. Note that c also has a neighbor $d_0 \in D$ such that $d_0 \neq d$, as otherwise c and d_0 could not be adjacent. Now, consider the central path $d_0 - c - v - d$, and all associated edges. This almost creates $S_{\kappa \times \kappa \times r \times \kappa}$, except we notice that c and d actually share a given vertex as an endpoint of an emanating edge. However, this is not a problem, since simply not considering this edge gives a $S_{\kappa-1 \times \kappa-1 \times r \times \kappa-1}$, which, since $\kappa > k$, implies our graph is necessarily saturated.

Finally, we need to construct this graph up to Q_n from Q_κ, in which case we simply place our saturated graph H at the vertices of $Q_\kappa \,\square\, Q_{n-\kappa}$. Note that we do not need to worry about respective vertices of our hamming codes being adjacent, because adding edges between them creates an $S_{k \times k \times r \times k}$ of the form $c_0 - c_1 - d_0 - d_1$, since each of these have degree at least $\kappa - 1 \geq k > r$.

From here, all that remains is to enumerate the number of edges, which as an upper bound gives

$$(2\kappa - 1) \cdot \frac{2^\kappa}{\kappa + 1} + (r - 3) \cdot \frac{\kappa 2^\kappa - 2^{\kappa+1}}{\kappa + 1} \leq (r - 1) \cdot 2^n.$$

Now, consider the second case, where $k_1 = \min\{k_1, k_2, k_3, k_4\}$ or $k_4 = \min\{k_1, k_2, k_3, k_4\}$. Again, both of these easily reduces to the $S_{k\times k\times k\times r}$ case given that the dominating set we construct has no 4 adjacent vertices with degree r or more. Therefore, start by considering the smallest $\kappa > k$ which satisfies $\kappa = 2^j - 1$ for some positive integer j. From here, we will attempt to find a $S_{k\times k\times k\times r}$-saturated graph within Q_κ.

Let us begin by denoting our saturated graph by H. Now, consider three dominating hamming codes, say C, $D = C + e_1$, and $E = C + e_2$. Each of these has minimum distance 3. From here, add all edges incident to the vertices in each of the hamming codes. Then, among the remaining vertices not in $C \cup D \cup E$, which at the moment form a 3-regular bipartite graph, find $r - 4$ perfect matchings using Lemma 4.14 and add these edges, thereby creating a subgraph H where all vertices in $C \cup D \cup E$ have degree κ, and all other vertices have degree $r - 1$.

From here, all that remains is showing that H is saturated in Q_κ, and then demonstrating that we can scale it up to Q_n. To do the former, we first note that it is clearly H-free, since there are no paths of length 3 among vertices with degree κ, since by construction our hamming codes form paths of length 2, and no more. Furthermore, if we add any non-edge to H, it must necessarily be incident to some vertex v with degree $r - 1$. Considering its neighbors $c \in C$, $d \in D$, $e \in E$, $v - d - c - e$ contains $S_{k\times k\times k\times r}$, so we are done. The only problem arises when the stars around e and d share some vertex, but since $\kappa > k$, we can simply not consider these edges and still have a $S_{k\times k\times k\times r}$. Therefore, H is saturated. To show that this construction can scale, simply place a copy of H at every vertex of $Q_\kappa \,\square\, Q_{n-\kappa}$.

Since connecting any two vertices in respective hamming codes creates a path of length 6 with all vertices with degree at least κ (sufficiently long for our 4-star), this subgraph of Q_n is $S_{k\times k\times k\times r}$-saturated.

All that remains is to enumerate the number of edges in our graph, which gives us an upper bound of

$$(3\kappa - 2)\cdot\frac{2^\kappa}{\kappa+1} + (r-4)\cdot\frac{\kappa 2^\kappa - 2^{\kappa+2}}{\kappa+1} \leq (r-1)\cdot 2^n.$$

$\square$

NOTE: In this proof, for the first case, we must have all $k_i \geq 2$, and, for the second case, we must have all $k_i \geq 3$. Otherwise, these are completely general results.

Proof of Theorem 4.18. Let $\max\{k_1, k_2, \ldots, k_m\} = k$, and $\max\{a_j, a_{j+1}\} = r$. We again use a κ-construction, picking $\kappa > a \cdot k$ for reasons that will become clear later. We will show that we can find a $S_{\lfloor\frac{\kappa}{a}\rfloor\times\lfloor\frac{\kappa}{a}\rfloor\times\cdots\times\lfloor\frac{\kappa}{a}\rfloor\times r\times r\times\lfloor\frac{\kappa}{a}\rfloor\times\cdots\times\lfloor\frac{\kappa}{a}\rfloor}$-saturated graph, which implies our result.

Let us begin by denoting our saturated subgraph by H. Now, let $j_0 = \max\{j, m-j\}$. We know that $j_0 < 2^a - 1$, so P_{j_0} can be embedded in Q_a. Therefore, consider a dominating set S consisting of disjoint Q_a's, with κ chosen accordingly so that this dominating set is perfect. Within our saturated graph H, we first add all edges within these Q_a's and incident to these Q_a's. Then, we use Lemma 4.14 on the remaining vertices to consistently add perfect matchings that give all other vertices degree $r-1$. However, we do this in such a way that every pair of adjacent vertices v and w in Q_n that are also adjacent to the same Q_a in S is connected by an edge. This requires that $r \geq a$, a condition already satisfied since $a = \lfloor \log_2 m \rfloor$. From here, we now have that every vertex in S has degree κ, and the remaining vertices in Q_n has degree $r-1$.

Given our construction of H, we claim that this graph H is saturated with our given star. To show this, first note that it is necessarily $S_{k_1\times k_2\times\cdots\times k_m}$-free, since the longest path in any given Q_a is $2^a - 1$, but $m \geq 2^a$. Now, consider adding any edge in Q_n to H. Since we have already connected all adjacent vertices adjacent to the same disjoint Q_a, this edge must connect adjacent vertices, say v and w, adjacent to different Q_a. In particular, both of these vertices must have had degree $r-1$ before, and now have degree r. Let a_x and a_y be the adjacent vertices to v and w in S. Now, consider a path of length $2^a - 3$ on the first hypercube ending at a_x, say $a_1 - a_2 - a_3 - \cdots - a_x$, and a path of length $b+1$ in the second hypercube starting at a_y. Since the longest path in Q_a is $2^a - 1$, this is evidently possible. We claim that, using these vertices and the edges surrounding them, it is possible to create $S_{\lfloor \frac{\kappa}{a} \rfloor \times \lfloor \frac{\kappa}{a} \rfloor \times \cdots \times \lfloor \frac{\kappa}{a} \rfloor \times r \times r \times \lfloor \frac{\kappa}{a} \rfloor \times \cdots \times \lfloor \frac{\kappa}{a} \rfloor}$. To create this, note, given any vertex adjacent to a vertex in our sequence, there could be at most $a-1$ other vertices in our sequence also adjacent to it. Since k-stars contain no cycles, we see now why κ necessarily must be greater than $a \cdot k$, which allows us to keep all such edges vertex-disjoint if we limit the number of edges emanating from each vertex to $\lfloor \frac{\kappa}{a} \rfloor$. Since $\frac{\kappa}{a} > k$, this applies equally to our original star. Note further that it is clear now why this implies that the graph is $S_{k_1 \times k_2 \times \cdots \times k_m}$-saturated, since all of the arguments about saturation apply equally to this k-star, given the minimum degree r is in the right place in the sequence.

The next step is to scale this graph up to Q_n. But, in this case, simply placing H at all vertices of $Q_\kappa \,\square\, Q_{n-\kappa}$ does the trick, since any non-edge now either connects two vertices with degree $r-1$ adjacent to different Q_a, or connects two Q_a. In this case, we can construct a path of length $2^a - 1 + 1 + 2^a - 1 = 2^{a+1} - 1 \geq m$ with all vertices having degree at least $\frac{\kappa}{a} > k$, so our original k-star is constructed. From here, all that remains to enumerate the number of edges in our saturated subgraph, which gives us, as desired, an upper bound of

$$2^{n-\kappa} \cdot \left(a \cdot \frac{2^{\kappa-1}}{\kappa - a + 1} + (\kappa - a) \cdot \frac{2^\kappa}{\kappa - a + 1} + (r-2) \cdot \frac{(\kappa - a + 1) \cdot 2^\kappa - 2^\kappa}{\kappa - a + 1} \right) \leq (r-1) \cdot 2^n.$$

$\square$

Proof of Theorem 4.19. Letting $\max\{k_1, k_2, \ldots, k_m\} = k$ and $\operatorname{emin}\{k_1, k_2, \ldots, k_m\} = r$, we use a similar construction as in Theorem 4.18, except we do not include edges between adjacent vertices adjacent to the same Q_a. It is still clear that any non-edge connects two vertices of degree $r-1$, forcing a path of length $2^a + 1$ to be formed, with the first $2^a - 1$ vertices having degree at least $\frac{\kappa}{a}$ and the final two having degree r. This evidently contains $S_{k\times k\times\cdots\times r\times r}$, as $\frac{\kappa}{a} > k$, so our graph is saturated. From here, since the construction is exactly the same except with fewer edges, the upper bound remains at $(r-1)\cdot 2^n$, so we are done. □

Appendix F: Beyond Caterpillars

In this final appendix, we present another, more general tree for which we can construct an upper bound. This particular class of trees is interesting because it is a sort of amalgamation between generalized stars and subdivided stars.

Our tree consists first of a central node with degree 3. Emanating from this central node, we have three paths of length 3, say $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$. Denote the nodes of any given path by $\mathcal{P}_{ik}$, where k ranges from 0 to 3 ($\mathcal{P}_{i0}$ simply represents the central node). In this tree, $\mathcal{P}_{i1}$ and $\mathcal{P}_{i2}$ have any degree, say k_{i1} and k_{i2}. Then, $\mathcal{P}_{i3}$ has three separate degrees depending on i: for $i = 1$, the degree is some integer $r > 1$, for $i = 2$ the degree is also r, and for $i = 3$, the degree is some integer $r_0 \geq r$. We also have the condition that $k_{ij} \geq r$ for all i, j. An example of this tree is shown in Figure 8.

For convenience, we will denote this class of trees by $\mathcal{T}$. The subclass within $\mathcal{T}$ whose P_{13} has degree r is denoted by $\mathcal{T}_r$. Now, we state our main result concerning this class of trees.

Theorem F.1. *Given some* $T \in \mathcal{T}_r$*,* $\operatorname{sat}(Q_n, T) \leq (3r + k_{21} - 4) \cdot 2^{n-2}$.

Proof. As usual, we make a κ-construction in Q_κ, this time with $\kappa = 2^i + 2$ and $\kappa > 3 \cdot \max\{k_{i1}, k_{i2}, r_0\}$. From this, we find a dominating set S of Q_3's in Q_κ, made possible by

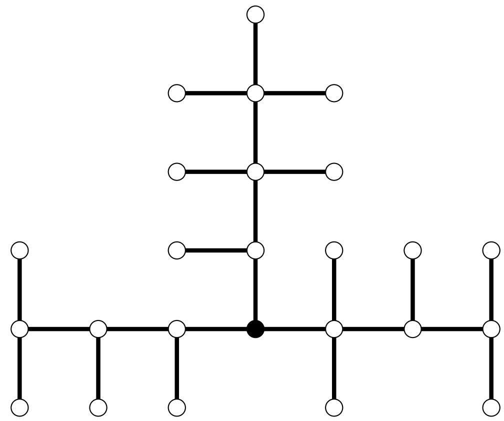

Figure 8: An example of the tree in question with $k_{11} = 4$, $k_{12} = 3$, $k_{22} = 3$, $k_{21} = 3$, $k_{31} = 3$, $k_{32} = 4$, $r = 2$, $r_0 = 3$. The black vertex represents the central node.

our choice of κ.

Now, denote our saturated subgraph by H. Begin our construction of H by adding all edges incident to and within our Q_3's. Then, among the remaining vertices, we use Lemma 4.14 to ensure that they all have degree $r-1$. Finally, within each Q_3 in S, pick two adjacent vertices and all of their κ adjacent vertices, and add an adjacent edge to each of these vertices, giving them degree r. This operation almost ensures saturation, except for non-edges between these vertices now with degree r. Without loss of generality assume $k_{11} \leq k_{21} \leq k_{31}$. To ensure that our graph is completely saturated, we add these non-edges to H in a symmetric fashion to make them all have degree $k_{21} - 1$. Notice that, in our construction, adjacent vertices on Q_3's with degree r do not create T because, as seen later in Figure 9, the endpoints of the paths emanating outwards must be both at the same parity in Q_3, and therefore cannot be adjacent.

We now claim that H is saturated. To see this, consider all possible non-edges. Note that a non-edge must either be incident to some vertex with degree $r-1$, or to two vertices both with degree $k_{11} - 1$. In the latter case, let these two vertices be v and x. v must also have another neighbor in Q_n with degree $k_{21} - 1$, say y, and also, evidently, a neighbor on some Q_3, say n. Therefore, taking v as the central node, we see that we can find a copy of T

by using paths in the adjacent Q_3's, all of which have vertices with degree κ, greater than any degree in T. There could potentially be intersections between stars that are exactly two apart, but because $\kappa > 3\max\{k_{i1}, k_{i2}, r_o\}$, it makes no difference in the creation of T. The only exception would occur if $k_{11} = k_{21}$, in which case we greedily add at most 1 edge from each of the $\frac{2^\kappa}{\kappa-2}$ vertices in consideration to maintain saturation.

Thus, the only case we now need to consider is when a non-edge is incident to a vertex with degree $r-1$, thereby giving it degree r. To address this case, consider Figure 9.

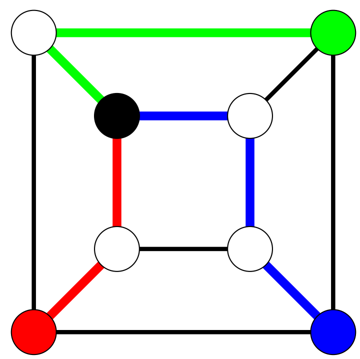

Figure 9: Paths in the cube, starting at the black vertex. We see that there exist vertex-disjoint paths of length 2, 2, and 3 respectively.

We will use this diagram to construct T with our non-edge added. Note that this Q_3 can be rotated in any way, so assume our vertex with degree r is adjacent to the red vertex. Follow the red path back to the originating black vertex. Each of these vertices has degree κ, which is greater than that of any $\mathcal{P}_{ik}$. Though there can be some intersections between our degree r vertex and the white vertex or the red vertex and the black vertex in this path, the number of edges able to be used is still clearly lower-bounded by $\frac{\kappa}{3}$, which is still sufficient given the definition of κ. A similar process on the blue path, which is our $\mathcal{P}_3$, shows that all vertices have sufficiently large degree. Finally, for our orange path, we claim that we can assume that this is incident to one of our degree $k_{21}-1$ vertices. We know that one of them must be the same parity as the degree r vertex adjacent to our red vertex, and we realize that we can simply alter our graph by rotation if it is any of the other vertices of the same parity. This ensures that the orange vertex is incident to a vertex with degree $k_{21} - 1 \geq r$, and thus, by similar arguments, the remaining vertices must have degree at least $\frac{\kappa}{3}$, sufficient to create that leg of our star. Thus, all three legs are created by the addition of any non-edge, but the initial subgraph does not contain T, meaning that our graph is saturated.

To scale this construction up to Q_n, notice that simply placing H at vertices of $Q_\kappa \square Q_{n-\kappa}$ creates a saturated graph. The only new complication introduced is edges between Q_3's, but it is easy to see, by a similar argument to above, that this also creates T using the orange and blue paths in our original hypercube along with any path in the corresponding hypercube in an adjacent H. From here, all that remains is to enumerate the number of edges in our subgraph, which gives us an upper bound of

$$2^{n-\kappa} \cdot \left(\kappa \cdot \frac{2^\kappa}{\kappa - 2} + (r-2) \cdot \frac{(\kappa-2)2^\kappa - 2^\kappa}{\kappa - 2} + (k_{21} - r) \cdot \frac{(\kappa - 2)2^{\kappa-2} - 2^{\kappa-2}}{\kappa - 2} \right) \leq (3r + k_{21} - 4) \cdot 2^{n-2},$$

as desired. $\square$

Note: this argument could be extended for central nodes with degree k if we could resolve, as in the generalized star case, the maximal length of k disjoint, equivalent paths in Q_κ. However, in general, this example shows how our argument using dominating sets in Q_k is viable for bounding the saturation number of perhaps any tree with small enough diameter relative to its minimum degree.

The second class of trees that we examine are a generalization of generalized stars. In particular, let an r-generalized star be the generalized star where every node in a leg is replaced by S_r. Denote this generalized star by $GS_{k,m,r}$. We find upper bounds on the saturation number for $GS_{k,2,r}$. Before presenting this proof, note that it actually generalizes so that only the final node in each leg needs to have degree r for the upper bound to apply.

Theorem F.2. *For all positive integers k, r that satisfy $k > r$,* $\operatorname{sat}(Q_n, GS_{k,2,r}) \leq (r-1)\cdot 2^n$.

Proof. Begin by constructing a κ-construction for some sufficiently large $\kappa = 2^i - k - 2$. By definition, this contains a dominating set S of Q_{k-1}'s. Let our saturated graph of Q_κ be denoted by H. Begin by adding all edges incident to and within S to H. Then, using Lemma 4.14, add $r-2$ perfect matchings to the remaining vertices such that, in the end, H contains only vertices with degree κ and with degree $r-1$. In particular, in this construction, we make sure to add all edges between vertices that are adjacent to the same Q_{k-1}.

We claim that this graph, in Q_κ, is $(GS_{k,2,r})$-saturated. First note that this is free of $GS_{k,2,r}$ because there exists some direction from each $s \in S$ for which the adjacent vertex has degree less than r. However, it is also saturated, because any incident edge in this graph clearly is between vertices with degree $r-1$. Also noting that, by construction, these are not adjacent to the same Q_{k-1}, we have a r-star path of length 2 adjacent to some vertex in a disjoint $Q_{k-1} \in S$. From this vertex, by Lemma 4.10, we also have $k-1$ paths of length 2 in our Q_{k-1}, all of whose nodes have degree at least $\frac{\kappa}{k}$, which, for sufficiently large κ, constructs $GS_{k,2,r}$. Thus, H is saturated.

To scale this to Q_n, we simply place H at all of the vertices of $Q_\kappa \,\square\, Q_{n-\kappa}$, as edges between hamming codes also evidently create $GS_{k,2,r}$. This, thus, maintains saturation and is still $GS_{k,2,r}$-free, since no additional edges were added between Q_κ's. Therefore, all that remains is to enumerate the number of edges in this subgraph, which gives us an upper bound of

$$2^{n-\kappa} \cdot \left(\kappa \cdot \frac{2^\kappa}{\kappa - k + 1} + (r-2) \cdot \frac{(\kappa - k + 1)2^\kappa - 2^\kappa}{\kappa - k + 1}\right) \leq (r-1) \cdot 2^n.$$

□