Bounded Degree Spanners of the Hypercube

Rajko Nenadov  Mehtaab Sawhney
ETH Zürich  Massachusetts Institute of Technology
Switzerland  Cambridge, U.S.A.
rajkon@gmail.com  msawhney@mit.edu

Benny Sudakov*  Adam Zsolt Wagner
ETH Zürich  ETH Zürich
Switzerland  Switzerland
{benjamin.sudakov,zsolt.wagner}@math.ethz.ch

Submitted: Oct 20, 2019; Accepted: Jun 15, 2020; Published: Jul 10, 2020
© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

In this short note we study two questions about the existence of subgraphs of the hypercube $Q_n$ with certain properties. The first question, due to Erdős–Hamburger–Pippert–Weakley, asks whether there exists a bounded degree subgraph of $Q_n$ which has diameter $n$. We answer this question by giving an explicit construction of such a subgraph with maximum degree at most 120.

The second problem concerns properties of $k$-additive spanners of the hypercube, that is, subgraphs of $Q_n$ in which the distance between any two vertices is at most $k$ larger than in $Q_n$. Denoting by $\Delta_{2k,\infty}(n)$ the minimum possible maximum degree of a $k$-additive spanner of $Q_n$, Arizumi–Hamburger–Kostochka showed that

$$\frac{n}{\ln n} e^{-4k} \leq \Delta_{2k,\infty}(n) \leq 20 \frac{n}{\ln n} \ln \ln n.$$

We improve their upper bound by showing that

$$\Delta_{2k,\infty}(n) \leq 10^{4k} \frac{n}{\ln n} \ln^{(k+1)} n,$$

where the last term denotes a $k + 1$-fold iterated logarithm.

Mathematics Subject Classifications: 05C88, 05C89

*Research supported in part by SNSF grant 200021-175573.
1 Introduction

Let $Q_n$ denote the hypercube graph, with vertex set $\{0, 1\}^n$ with edges connecting two vertices if they differ in precisely one coordinate. Sparse subgraphs of the hypercube with strong distance-preserving properties have been studied extensively in the literature, and have found many practical applications in distributed computing and communication networks. We refer the reader to the recent survey [2].

Erdős–Hamburger–Pippert–Weakley [16] studied spanning subgraphs of $Q_n$ with diameter $n$. They observed that there exists such a subgraph with average degree $2 + O\left(\frac{1}{\sqrt{n}}\right)$, however in their construction there were vertices of degree $n$. They asked the following natural question:

**Question 1** (Erdős–Hamburger–Pippert–Weakley [16]). Does there exist a spanning subgraph of $Q_n$ with bounded degree and diameter $n$?

Our first result is an explicit construction giving a positive answer to Question 1.

**Theorem 2.** There exists a spanning subgraph $G$ of $Q_n$ with maximum degree at most 120 such that the diameter of $G$ is $n$.

One particular distance-preserving property that has received much attention in the past is that of an additive spanner. We say a subgraph $G \subset Q_n$ is a $k$-additive spanner if $\text{dist}_G(x, y) \leq \text{dist}_{Q_n}(x, y) + k$ for any two vertices $x, y \in \{0, 1\}^n$.

Constructions of additive spanners with few edges and/or low maximum degree have attracted considerable attention in computer science in the past. A 2-additive spanner of $Q_n$ with average degree $\frac{n+1}{2}$ can be found in [18]. The first 2-additive spanner constructions for general graphs are given in [3]; 4- and 6-additive spanners are found in [13, 9]. A barrier to any further general additive spanner constructions is discussed in [1]. Papers specifically dealing with the low maximum degree case are e.g. [11, 15, 14, 17].

Arizumi–Hamburger–Kostochka [4] denoted by $\Delta_{k,\infty}(n)$ the minimum possible maximum degree of a $k$-additive spanner of $Q_n$. Note that since $Q_n$ is bipartite, by deleting edges the distance can only grow by an even amount. They showed that for $k \geq 2$ and $n \geq 21$ we have

$$\frac{n}{\ln n} e^{-4k} \leq \Delta_{2k,\infty}(n) \leq 20 \frac{n}{\ln n} \ln \ln n.$$  

Their lower bound is a short argument given by counting the vertices of a certain distance from a fixed vertex, and their upper bound is an explicit construction. Our second result is an improvement of their upper bound on this problem.

**Theorem 3.** For all $n$ sufficiently large and $k \in \mathbb{N}$, there exists a $2k$-additive spanner of the hypercube with maximum degree at most

$$10^{k} \frac{n \ln^{(k+1)} n}{\ln n}.$$
Note here that $\ln^{(k+1)} n$ is the $k + 1$-times iterated logarithm, defined by $\ln^{(1)} n = \ln n$ and $\ln^{j+1} n = \ln \left( \ln^{j} n \right)$.

We prove Theorem 2 in Section 2 and prove Theorem 3 in Section 3. Some open questions and further directions of study are given in Section 4.

2 Bounded degree subgraph preserving diameter

In the present paper, a perfect code will always mean a perfect 1-error-correcting code over the alphabet \{0, 1\} with codewords having length $n$. We say that $C$ is a perfect 1-error-correcting code if any two codewords have Hamming distance at least three, and moreover the radius one Hamming balls centered on the codewords partition the whole space \{0, 1\}^n. Note that the number of codewords in a perfect code $C$ is $|C| = 2^n/n + 1$. Perfect codes exist whenever $n = 2^r - 1$ for some $r \in \mathbb{N}$, see e.g. [19]. We will use the fact that for all $n = 2^r - 1$, $r \in \mathbb{N}$, it is possible to partition the space \{0, 1\}^n into $n + 1$ perfect codes (see e.g. [19], p. 15.).

In this section we prove Theorem 2. We first need the following technical lemma.

**Lemma 4.** For all $n$ there is a subset $S$ of vertices of $Q_n$ with the following two properties.

- Every vertex $v$ is either in $S$ or adjacent to a member of $S$.
- Every vertex $v$ is adjacent to at most 2 vertices in $S$.

We refer to such a subset of vertices $S$ of $Q_n$ as a nearly perfect code.

**Proof.** Note that for $n = 2^k - 1$ the result follows from the existence of perfect codes. For other values of $n$ let $k$ be such that $n$ is between $2^k - 1$ and $2^{k+1} - 1$ and divide the $n$ coordinates into $2^k - 1$ buckets of size at most 2. Now take a perfect code $C$ of $Q_{2^k-1}$. We define $S$ as follows. Given an element $v \in Q_n$, for all $1 \leq i \leq 2^k - 1$ define $b_i$ to be the sum of the elements of $v$ in the $i$-th bucket, where this sum is taken over $\mathbb{F}_2$. Then the element $v$ is in $S$ if and only if the element $b_1b_2\ldots b_{2^k-1}$ is in $C$. It is now straightforward to verify that $S$ satisfies the necessary conditions. Indeed, given any $w \in Q_n$, consider the word $u \in Q_{2^k-1}$ obtained from by by taking the sum in $\mathbb{F}_2$ of each bucket. Then $u$ is either in $C$ or adjacent to some $u'$ in $C$. In the first case $w \in S$ and $w$ is not adjacent to any other element of $S$. In the second case $w$ is adjacent to exactly all those $w' \in C$ which agree with $w$ everywhere except on the block which corresponds to the bit in which $u$ and $u'$ differ. Moreover on this block $w'$ has different parity from $w$. Since all blocks have length at most 2, we conclude that $w$ is adjacent to at most two members of $S$. \qed

We now proceed first by proving a weaker version of Theorem 2 which only requires preserving the distance between antipodes. We say two vertices are antipodal if they differ on all $n$ coordinates.

**Lemma 5.** There exists a subgraph $H$ of $Q_n$ with max degree 10 such that all antipodes are at distance $n$ within $H$.

THE ELECTRONIC JOURNAL OF COMBINATORICS 27(3) (2020), #P3.3
Proof. Consider the set of vertices \((0, \ldots, 0), (1, 0, \ldots, 0), (1, 1, 0, \ldots, 0), \ldots, (1, \ldots, 1), (0, 1, \ldots, 1), (0, 0, 1, \ldots, 1), \ldots, (0, \ldots, 0, 1)\), i.e. all vertices with coordinates having all 0’s and then followed by 1’s or having all 1’s and then followed by 0’s. Note that these points form a cycle \(C\) of length \(2n\) in \(Q_2\) such that there are \(n\) pairs of antipodes along this cycle. In particular, if a vertex is on \(C\) then there is a path of length \(n\) from this vertex to its antipode using only edges of \(C\).

The construction of \(H\) is to translate this \(2n\)-cycle \(C\) by the appropriate nearly perfect code so that every vertex is contained in one of these cycles. By the above discussion, if some vertex is on a translation of \(C\) then so is its antipode and therefore their distance in \(H\) is \(n\).

Let \((e_i)_{i=1}^n\) be the standard basis vectors in \(\mathbb{F}_2^n\) and for \(1 \leq k \leq n\) define \(f_k\) as

\[ f_k = \sum_{i=1}^{k} e_i. \]

Note that the vectors \(f_i\) form a basis of \(\mathbb{F}_2^n\) and that the vertices of the cycle \(C\) are exactly \(0, f_1, \ldots, f_n, f_n - f_1, f_{n-2}, \ldots, f_n - f_{n-1}\). Now consider a nearly perfect code \(S'\) in the basis of the \(f_i\) vectors and all the translations \(s + C\) where \(s \in S'\). First note that every element is contained in at least one cycle as we have translated 0 by a nearly perfect code in the \(f_i\) basis and all basis vectors \(f_1, \ldots, f_n\) are in the cycle. To see that no vertex however is in more than five cycles consider the element \(s \in S'\) so that \(v \in s + C\). Then it follows that \(v = s, s + f_j, or s + f_n - f_j\) for some \(j\) and therefore \(s = v, v + f_j, or v - f_n + f_j\). Thus \(s\) is either \(v\), a neighbor of \(v\), or a neighbor of \(v - f_n\) and by the definition of nearly perfect codes we have that every vertex is either in the code or next to at most two other code words we have that every vertex is in at most 5 cycles and hence has degree at most 10.

We now use Lemma 5 to complete the proof of Theorem 2; note that the fact that the construction comes from a union of these antipodal cycles plays a critical role in Theorem 2.

Proof. For \(n \leq 100\) note that taking \(Q_n\) suffices. Otherwise let \(B_1, B_2, B_3, B_4\) be an equitable partition of the coordinates. That is, define \(n_i\) such that \(n = \sum_{i=1}^{4} n_i\) or \(n_i = \left\lceil \frac{n}{4} \right\rceil\) and \(n_i \geq n_j\) for \(i < j\). Now associate \(Q_n\) with its representation on \(\{0, 1\}^n\) and divide the coordinates in blocks \(B_1, B_2, B_3, B_4\) with \(B_i\) having block size \(n_i\). Finally for each vertex \(v\) define \(v_i\) to be restriction of \(v\) to the block \(B_i\).

Consider \(H_i\) that is a subgraph of \(Q_n\), coming from Lemma 5. Now consider an almost equal partition, \(M_i\), of \([n] \setminus B_i\) into \(n_i\) parts. As \(n \geq 100\), note that every part has size at most four. For each antipodal cycle \(C\) of \(H_i\) do the following. First, we pick an arbitrary order of the vertices of \(C\), say they are \(x_1, x_2, \ldots, x_{2n}\). Next, for each \(1 \leq i \leq n_i\), assign to vertex \(x_i\), the \(i\)-th part of \(M_i\). Note that at the end of this process, since vertices may be in multiple cycles in \(H_i\), some vertices may have multiple parts of \(M_i\) assigned to them.

We are now ready to define the neighbors of \(v\), of which there will be two types. The first set of neighbours will depend only on the \(H_i\)-s. For each \(1 \leq i \leq 4\), we let a vertex \(w\)
be a neighbour of \(v\) if they are equal on every coordinate outside of \(B_i\), and the restriction of \(w\) to the coordinates in \(B_i\) is adjacent to \(v_i\) in the graph \(H_i\). That is, for each \(1 \leq i \leq 4\) we keep the neighbors in the directions which \(v_i\) has in \(H_i\).

The second set of neighbours of \(v\) will depend on the partitions \(M_i\). For each \(1 \leq i \leq 4\) let \(P_i\) be the part of \(M_i\) to which \(v_i\) was assigned, and keep all edges in direction \(P_i\) — that is, let \(w\) be a neighbour of \(v\) if they are equal everywhere outside of the at most four coordinates of \(P_i\). This is the desired subgraph \(G\) of the hypercube.

We first verify that the max degree of \(G\) is at most 120. First note that since the \(H_i\) have max degree at most 10 and we have four parts the contribution to each vertex \(v\)'s degree from the \(H_i\) is bounded by 40. Furthermore since each part of \(M_i\) has size at most four, there are at most four edges outgoing from \(v\) due to this part. Since every vertex was on at most five cycles in \(H_i\), the total number of edges outgoing from \(v\) due to the partition of \([n]\setminus B_i\) into \(n_i\) parts is at most \(5 \cdot 4 = 20\). The only subtlety is now to account for edges incoming to \(v\) from each of the partitions of \([n]\setminus B_i\). But note that for such edges \(v\) to \(w\) we have \(v_i = w_i\) and therefore this relationship is in fact symmetric unlike the asymmetric description. That is, if in the above definition we have defined \(w\) to be in a neighbour of \(v\), then when considering the vertex \(w\) instead of \(v\) we have defined \(v\) to be a neighbour of \(w\) (of the same type). Therefore the total degree count is \(40 + 4(20) = 120\).

Finally we demonstrate that \(G\) is diameter \(n\). Consider two vertices \(v\) and \(w\) such that \(v_i\) and \(w_i\) match on exactly \(k_i\) coordinates. Suppose that \(k_1 \leq k_2 \leq k_3 \leq k_4\); the other cases are handled in an analogous manner. In order to “fix” \(v\) we first proceed along the cycle in \(H_1\) to \(v_1\)'s antipode in order to obtain \(x\). Along the way we can adjust the coordinates in \(B_3\) and \(B_4\) so that \(x_3 = w_3\) and \(x_4 = w_4\) while \(x_2\) is the antipode of \(w_2\). Note that we take \(n_1\) steps along \(H_1\), \(k_2\) to change the coordinates in \(B_2\) appropriately, and \(n_3 - k_3 + n_4 - k_4\) steps to fix \(B_3\) and \(B_4\). We now walk from \(x_2\) to \(w_2\) along \(H_2\) and along the way fix \(H_1\) so that it now matches \(x_1\). This takes \(k_1\) steps to fix \(B_1\) to \(w_1\) and \(n_2\) to fix \(B_2\) to \(w_2\). Therefore the total number of steps is

\[
n_1 + k_2 + n_3 - k_3 + n_4 - k_4 + k_1 + n_2
\]

\[
= (n_1 + n_2 + n_3 + n_4) + (k_1 + k_2 - k_3 - k_4)
\]

\[
\leq (n_1 + n_2 + n_3 + n_4) = n
\]

as desired. \(\square\)

### 3 \(k\)-additive Spanner of the Hypercube

The main goal of this section is to prove Theorem 3 by constructing a \(2k\)-additive spanner of the hypercube with small maximum degree. For the sake of clarity various floor and ceiling symbols will be omitted. The key idea is to essentially iterate the construction in [4] which achieves this result for the \(k = 1\) case (with a slightly better constant).

Before we begin let us first give a high level overview of the construction. The desired \(2k\)-additive spanner \(H\) will be the union of three subgraphs \(H_1, H_2, H_3\) that each play a different role in preserving distances.
The first subgraph, \( H_1 \), will simply connect two vertices if they differ in the first approximately \( \sqrt{n} \) coordinates. This graph has negligible maximum degree, and its purpose will be twofold. First, it enables us to prove the spanning property only for pairs of vertices which agree on the first \( \sqrt{n} \) coordinates, thus simplifying the problem. Second, we will partition the cube on the first \( \sqrt{n} \) coordinates into perfect codes, and \( H_1 \) will let us step into any of these codes with at most one move. Then in order to find a short path between two vertices \( x, y \), we first step from \( x \) into the appropriate code in the first \( \sqrt{n} \) coordinates, then find a path of length \( d_{Q_n}(x, y) + 2(k-1) \) that changes \( x \) to \( y \) everywhere outside of the first \( \sqrt{n} \) coordinates using \( H_2 \) and \( H_3 \), and then use \( H_1 \) to step from the code to \( y \). This will then give an \( x-y \) path of length \( d_{Q_n}(x, y) + 2k \).

The subgraph \( H_2 \) will be primarily used to find short paths between vertices that are not too far apart. The way we do this is roughly as follows. As mentioned before, we will partition the first \( \sqrt{n} \) coordinates into perfect codes. Each of these perfect codes will be responsible for a different small subset of the remaining \( n - \sqrt{n} \) coordinates. So if the vertices \( x, y \) differ in a small set \( S \) of the last \( n - \sqrt{n} \) coordinates then we can first move from \( x \) to a point \( x' \) using \( H_1 \), where \( x' \) belongs to the perfect code that is responsible for the set \( S \). Then \( H_2 \) will be defined so that we can fix the coordinates in \( S \) by losing at most \( 2(k-1) \) distance. This we can achieve by using a \( 2(k-1) \)-additive spanner on the set \( S \), which is given by induction on \( k \).

This leaves us with the task of finding short paths between vertices that are far apart. It is not possible to have a bijection between perfect codes on the first \( \sqrt{n} \) coordinates and all subsets of the remaining \( n - \sqrt{n} \) coordinates, as the number of possibilities of these subsets is too large. Instead we do the following. Given two vertices \( x \) and \( y \), we first find a small set \( S \) that contains many of the differences between these two vertices. We fix the differences in \( S \) using the graphs \( H_1 \) and \( H_2 \) as explained above. However, while walking in \( H_2 \), we will need to fix all the coordinates outside of \( S \) as well. This is where \( H_3 \) will come in. In order for this construction to work we will need to make our induction hypothesis stronger, and additionally require that the additive spanners we construct (and hence \( H_2 \) too) have the property that between any two points there is a short path that hits many different coordinate sums — that is, if we walk along this path and at every vertex we note down the sum of its coordinates, we want to see a large number of different values by the time we reach the end of the path.

The idea of \( H_3 \) is then similar to how we constructed \( H_2 \). Given two vertices \( x, y \) we find a small set \( S \) which contains many of the coordinates where they differ. Then we start walking along \( H_1 \) and \( H_2 \) to fix the set \( S \). While doing so, we keep track of the coordinate sum inside the set \( S \). By induction hypothesis, we will see many different sums. Each sum will correspond to a different small subset of \( [n] \setminus S \). So while walking in \( H_2 \), whenever we encounter a coordinate sum that corresponds to a set of coordinates which contains at least one difference between \( x \) and \( y \), we use \( H_3 \) to fix those coordinates. In \( H_3 \), we simply include all edges necessary to make these patches work. The fact that we see many different coordinate sums throughout our walk in \( H_2 \) ensures that we will get a chance to fix all coordinates outside of \( S \).

We are now ready to begin the proof and make the above sketch precise.
Proof. We build the construction iteratively as $k$ increases. We will maintain the following invariant for the $2k$-additive spanners: for any pair of points which are distance $\ell$ apart there is a path connecting them of length at most $\ell + 2k$ whose vertices have at least $\frac{\ell}{2^k}$ different coordinate sums. For $k = 0$ the construction is taking the entire hypercube graph $Q_n$. This satisfies the necessary maximum degree condition and will serve as the base case for this induction on $k$. Note that in $Q_n$ between any two points at distance $\ell$ we may first flip all necessary zero coordinates to ones and then all required ones to zeros, ensuring the existence of a path of length $\ell$ whose vertices have at least $\ell/2$ different coordinate sums. For the remainder of the proof we divide the coordinates into two groups.

- Pick $r \in \mathbb{N}$ so that $2^r - 1 \in [\sqrt{n}/2, \sqrt{n}]$, and let $B_0$ be the first $q = 2^r - 1$ coordinates.
- $B_1, \ldots, B_t$ will be an almost equal partition of the remaining $n - q$ coordinates into $t = \frac{\ln n \ln^{(t)} n}{900 (\ln^{(k+1)} n)^2}$ blocks.

We define an additional parameter $s$ to be $s = \frac{\ln n}{10 \ln^{(k+1)} n}$. The construction now has three distinct parts.

- Define $H_1$ to be subgraph created by including all edges in directions in $B_0$ for every vertex in $Q_n$. That is, two adjacent vertices $x, y \in Q_n$ are connected by an edge in $H_1$ precisely if they differ in only one of the first $q$ coordinates and nowhere else. Note that $H_1$ is a vertex-disjoint union of $2^{n-q}$ copies of $Q_q$.

- We now define the subgraph $H_2$. Since $q = 2^r - 1$, as remarked at the beginning of Section 2, we can partition each disjoint copy of $Q_q$ in $H_1$ into perfect codes $D'_1, \ldots, D'_{q+1}$ and let $D_i$ be the union of the $D'_i$ over these disjoint components of $H_1$. Now fix a bijective map $f$ from $\{D_1, D_2, \ldots, D\binom{\ell}{q}\}$ to $\binom{[\ell]}{q}$. Define the subgraph $H_2$ as follows. For every vertex $x$, first find the index $i$ so that $x$ belongs to $D_i$. If $i \leq \binom{\ell}{q}$ then let $\{B_{i_1}, B_{i_2}, \ldots, B_{i_q}\}$ be the set of $s$ blocks of coordinates from $B_1, \ldots, B_t$ corresponding to $D_i$. Next, let $B_x := B_{i_1} \cup \ldots \cup B_{i_q}$, fix all coordinates of $x$ on $[n] \setminus B_x$, and on the coordinates in $B_x$ include the $2(k-1)$-additive spanner on $|B_x|$ many coordinates that is given by the induction hypothesis. For example, if $x$ belongs to $B_3$ and $f(D_3) = \{1, 2, \ldots, s\}$ then we fix the coordinates of $x$ outside of $B_1 \cup \ldots \cup B_s$ and include the $2(k-1)$-spanner construction given by the induction hypothesis on the approximately $(n-q)s/t$ coordinates in $B_1 \cup \ldots \cup B_s$.

- We now define $H_3$. First divide $B_1, \ldots, B_t$ into groups of size $j = \frac{500k}{s}$; label these $A_1, \ldots, A_s/500k$. Now for each vertex $x$ which belongs to a $D_i$ with $i \leq \binom{\ell}{q}$, let $B_x := B_{i_1} \cup \ldots \cup B_{i_q}$ where $\{i_1, \ldots, i_q\} = f(D_i)$ as above. Next, let $s_x := \sum_{i \in B_x} x_i$. Take this coordinate sum $s_x \mod s/500k$, call this $s'$, and for the vertex $x$ only include edges in directions $A_{s'} \setminus B_{i_1} \cup \ldots \cup B_{i_q}$.

Note that in $H_2$ and $H_3$ the edges are (implicitly) “directed” from one vertex to another. However one can verify that the definitions are symmetric in both cases. First consider $H_2$. Note that if a vertex $y$ agrees with $x$ on every coordinate outside of $B_x$ then in
particular they agree on the first \( q \) coordinates used to define the sets \( D_i \). Therefore we have that \( x \) and \( y \) belong to the same \( D_i \) and thus \( B_x = B_y \). Hence in the construction of \( H_2 \) when we consider \( y \) we include the same \( 2(k - 1) \)-spanner construction on \( B_x \) as we did when considering the vertex \( x \).

Next consider \( H_3 \). If \( x \) is connected to \( y \) in \( H_3 \) this implies that the coordinate on which \( y \) differ from \( x \) lies outside of \( B_x \) and also outside of the set of the first \( q \) coordinates. Therefore, both \( x \) and \( y \) belong to the same set \( D_i \), \( B_x = B_y \), and also \( s_x = s_y \). So in both cases the value of \( s' \) is the same, hence in the construction of \( H_3 \) we included edges touching \( x \) and \( y \) in precisely the same directions.

The desired subgraph of \( Q_n \) is simply \( H = H_1 \cup H_2 \cup H_3 \). We first verify that the subgraph \( H_2 \) is well defined in that the desired bijection \( f \) indeed exists.

**Lemma 6.** For \( n \) sufficiently large, we have

\[
\binom{\ell}{s} \leq \sqrt{n}/2.
\]

**Proof.** Note that \( \frac{t}{s} = \frac{\ln(n)}{90 \ln^{k+1}(n)} \), in particular \( \frac{t}{s} \leq \ln(k)(n) \). Therefore

\[
\binom{\ell}{s} \leq \left( \frac{te}{s} \right)^s = \text{exp} \left( s \left( \ln \left( \frac{te}{s} \right) \right) \right) \leq \text{exp} \left( \frac{\ln(n)}{10 \ln^{k+1}(n) \ln(\ln(k)(n))} \right) \leq \sqrt{n}/2
\]

for \( n \) sufficiently large as desired. Note that in the last inequality we used that \( \ln^{k+1}(n) = \ln(\ln(k)(n)) \). \( \Box \)

We now prove the desired bound on the maximum degree of the graph \( H = H_1 \cup H_2 \cup H_3 \).

**Lemma 7.** The maximum degree of \( H \) is at most

\[
10^4 n \ln(\ln(k+1)(n)) \frac{n \ln(n)}{\ln(n)}
\]

for all \( k \) and for \( n \) sufficiently large.

**Proof.** We proceed by induction on \( k \). Note that for \( k = 0 \) this statement is trivial. For larger \( k \) note that the maximum degree in \( H_1 \) is at most \( \sqrt{n} \), the maximum degree in \( H_2 \) is upper bounded, using the induction hypothesis for the \( 2(k - 1) \)-spanner on \((n - q) \cdot \frac{s}{t}\) coordinates, by

\[
\frac{10^4(k-1) (n \cdot \frac{s}{t}) \ln^{(k)}(n \cdot \frac{s}{t})}{\ln(n \cdot \frac{s}{t})}
\]

The maximum degree in \( H_3 \) is at most \( \frac{500^k n}{s} \) for \( n \) sufficiently large. Now note that \( \frac{s}{t} = 90 \ln^{(k+1)}(n) \) and so

\[
\sqrt{n} + \frac{10^4(k-1) (n \cdot \frac{s}{t}) \ln^{(k)}(n \cdot \frac{s}{t})}{\ln(n \cdot \frac{s}{t})} + \frac{500^k n}{s} \leq 10^4 n \ln(\ln(k+1)(n)) \frac{n \ln(n)}{\ln(n)}
\]

for \( n \) sufficiently large as desired. \( \Box \)

---

**THE ELECTRONIC JOURNAL OF COMBINATORICS 27(3) (2020), #P3.3 8**
We first prove that it suffices to consider pairs of vertices whose coordinates match along $B_0$ due the presence of the subgraph $H_1$.

**Lemma 8.** If for all pairs of vertices $x, y$ that are equal on all coordinates in $B_0$ we have

$$d_H(x, y) \leq d_{Q_n}(x, y) + 2k$$

then the same follows for all pairs of vertices $x$ and $y$.

**Proof.** Consider an arbitrary pair of vertices $x$ and $y$. Let $x'$ such that the first $|B_0|$ coordinates of $x'$ match $y$ and the rest match $x$. Note that $d_H(x, x') = d_{Q_n}(x, x')$ and that $x'$ and $y$ satisfy the condition of the hypothesis. Therefore

$$d_H(x, y) \leq d_H(x, x') + d_H(x', y) \leq d_{Q_n}(x, x') + d_{Q_n}(x', y) + 2k = d_{Q_n}(x, y) + 2k$$

and the result follows. \qed

We now finally prove that $H$ is a $2k$-additive spanner. Recall that we are maintaining the invariant, that for any pair of points which are distance $\ell$ apart there is a path whose length is at most $\ell + 2k$ and whose points have at least $\ell/2$ different coordinate sums.

Furthermore note that the previous lemma does not interfere with this invariant; if the initial points differed in more than $\ell/2$ coordinates in $B_0$ then by applying the above mentioned procedure of first flipping 0 to 1 and then 1 to 0 we get a path with whose vertices give at least $\ell/2 = \ell/4$ different coordinate sums which is sufficient. Otherwise, the points differ by at least $\ell/2$ coordinates outside of $B_0$. Therefore it is enough to maintain the invariant that among points with coordinates that match in $|B_0|$ and are of distance $\ell$, there is path of whose vertices give $\ell/16$ coordinate sums. We now consider two cases.

**Case 1:** Suppose that $x$ and $y$ differ on at most $s$ coordinates. Then there exist \(\{i_1, \ldots, i_s\}\) such that the set of differences is contained inside $B_{i_1}, \ldots, B_{i_s}$. Let $D := f^{-1}(\{i_1, \ldots, i_s\})$. First we go from $x$ and $y$ to the closest points, call these points $x'$ and $y'$ respectively, in the perfect code $D$ using edges in $H_1$. Note that since $x$ and $y$ agree in the coordinates in $B_0$, so do $x'$ and $y'$. Now walk from $x'$ to $y'$ using edges of the $2(k - 1)$-spanner used to construct $H_2$ and adjust the necessary bits in $B_{i_1}, \ldots, B_{i_s}$. This can be done because $x'$ and $y'$ only differ in coordinates in $B_{i_1}, \ldots, B_{i_s}$ and by the definition of $2(k - 1)$ spanner the length of this path is by at most $2(k - 1)$ larger than their distance in $Q_n$. Therefore, the length of the path that we construct from $x$ to $y$ is at most $d_{Q_n}(x, y) + 2(k - 1) + 2 = d_{Q_n}(x, y) + 2k$. Moreover we can maintain the invariant regarding coordinate sums by invoking the inductive hypothesis and noticing that $d_{Q_n}(x, y) = d_{Q_n}(x', y')$.

**Case 2:** Suppose that $x$ and $y$ differ on more than $s$, say $\ell$, coordinates. Then find sets $B_{i_1}, \ldots, B_{i_s}$ such that $x$ and $y$ differ by at least $s$ coordinates in $B_{i_1} \cup \ldots \cup B_{i_s}$. We consider two separate situations.

- Suppose that $B_{i_1}, \ldots, B_{i_s}$ contain more than $\ell/5$ of the coordinate differences. In this case again we first move from $x$ and $y$ to the closest points $x'$ and $y'$ (respectively),

---

**THE ELECTRONIC JOURNAL OF COMBINATORICS 27(3) (2020), #P3.3**
in the perfect code $D = f^{-1}(i_1, \ldots, i_s)$ using edges in $H_1$. Then we use the edges in $H_2$ and the path given by the inductive hypothesis to change coordinates of $x'$ so that it agrees with $y'$ on $B_{i_1} \cup \ldots \cup B_{i_s}$. However when we visit the vertices which give us the first $\frac{\ell}{10 \cdot 32^k}$ coordinate sums out of the $\frac{s}{500^k}$ which is guaranteed by the inductive hypothesis, we keep them to satisfy the inductive hypothesis regarding coordinate sums, and we use the vertices with the remaining $\frac{\ell}{10 \cdot 32^k}$ coordinate sums to access all the blocks outside of $B_{i_1} \cup \ldots \cup B_{i_s}$ and use the subgraph $H_3$ to fix the differences between $x'$ and $y'$. More precisely, every time we take a step using $H_2$ in a direction in $B_{i_1} \cup \ldots \cup B_{i_s}$ to reach the point whose sum of the coordinates gives us a new residue modulo $s/500^k$ we check if there are any coordinates we can fix using $H_3$, and fix them. As $\frac{\ell}{10 \cdot 32^k} \geq \frac{s}{500^k}$ we will see all the residues modulo $s/500^k$ and so eventually we will be able to fix all the blocks using $H_3$. Note here that we only lost distance 2 in moving to $x'$ and $y'$ and distance $2(k-1)$ in making $x'$ equal to $y'$ on $B_{i_1} \cup \ldots \cup B_{i_s}$. We did not lose any distance anywhere else.

• Finally suppose that $B_{i_1}, \ldots, B_{i_s}$ contains less than $\frac{\ell}{5}$ of the differences. In this case we have that $x$ and $y$ differ on at least $4\ell/5$ coordinates outside of $B_{i_1} \cup \ldots \cup B_{i_s}$. Again, we first move into the points $x'$ and $y'$ in the perfect code $D = f^{-1}(i_1, \ldots, i_s)$ and start fixing the coordinates in $B_{i_1} \cup \ldots \cup B_{i_s}$ using $H_2$ and the path given by the inductive hypothesis. Since $x$ and $y$ differ on at least $s$ coordinates in $B_{i_1} \cup \ldots \cup B_{i_s}$, the vertices of this path have at least $s/32^k$ different coordinate sums. We fix the coordinates outside of $B_{i_1} \cup \ldots \cup B_{i_s}$ in a similar way as in the previous case: after every step we take in $H_2$ we try to fix as many coordinates as possible using edges in $H_3$. However the one difference in this case is, that during the stretch of $\frac{s}{32^k} \geq 2 \cdot \frac{s}{500^k}$ different coordinate sums we encounter while walking in $H_2$, we use the first half of them to fix all those coordinates where $x$ is zero and $y$ is one, and we use the second half to map all 1 to 0. At least one of these halves has length greater than $\frac{2\ell}{5}$, let us assume without loss of generality that it is the half where we change the zeros in $x$ to ones. During these at least $\frac{2\ell}{5}$ steps, the sum of coordinates increases by $+1$ each time. However, we have no control over the steps that we take in $H_2$, they can both increase and decrease the coordinate sum by 1. We have taken at most $\frac{\ell}{5} + 2(k-1)$ steps in $H_2$, so this part of the path has to give at least $\frac{2\ell}{5} - (\frac{\ell}{5} + 2(k-1)) \geq \frac{\ell}{10}$ different coordinate sums. This allows us to maintain the desired invariant and walk between $x$ and $y$ in the required length. \qed

4 Concluding remarks and open questions

The effect of the removal of a set of vertices or edges from computer networks, corresponding to broken connections, processors or inaccessible agents, are of major interest in the study of vulnerability of networks. Parameters that measure changes given by such breakdowns lead to many interesting open problems [5, 6], from which we only mention a few here.
The integrity $I(Q_n)$ of the hypercube is defined as

$$I(Q_n) = \min\{|S| + m(Q_n \setminus S) : S \subset V(Q_n)\},$$

where $m(H)$ denotes the number of vertices in the largest connected component of $H$. It is known (see [8, 10]) that

$$c_2^n \sqrt{n} \leq I(Q_n) \leq C_2^n \sqrt{n \log n},$$

and determining the precise asymptotics would be of interest. Another important related concept is that of “fault-tolerant spanners” (see e.g. [12, 20]) where the aim is to find spanners that retain their metric properties even after a number of components fail. For example one might want to find a subgraph $H$ of $Q_n$ with the property that for any collection $F$ of constantly many failing edges $F$ and any two vertices $x, y \in Q_n \setminus F$ we have $d_{H \setminus F}(x, y) \leq d_{Q_n \setminus F}(x, y) + 2k$.

The second problem we have already hinted at in the introduction.

**Question 9** (Erdős–Hamburger–Pippert–Weakley [16]). What is the least possible number of edges in a graph $G \subset Q_n$ that has diameter $n$?

They observed that there is such a graph $G \subset Q_n$ with $2^n + \left\lfloor \frac{n}{2} \right\rfloor - 2$ edges. The construction is as follows: from the all zero vertex and the all one vertex take two BFS trees until the middle layer; if $n$ is odd then there are two middle layers, and we add a perfect matching between them. We have the following lower bound on Question 9:

**Proposition 10.** If $G \subset Q_n$ has diameter $n$ then $e(G) \geq 2^n + \Theta\left(\frac{2^n}{n}\right)$.

**Proof.** Let $G \subset Q_n$ be a subgraph with diameter $n$. First we show that $G$ has minimum degree at least 2. Indeed assume $v$ is a leaf, connected only to vertex $u$. Let $u'$ denote the antipodal vertex to $u$. Then the distance of $v$ and $u'$ in $G$ is at least $n + 1$, contradicting the fact that $G$ has diameter $n$.

Next, fix an arbitrary vertex $v$ and consider a BFS tree $T$ rooted at $v$. Since $G$ has diameter $n$, the tree $T$ has at most $n + 1$ layers (the first layer being the single vertex $v$) and hence has a layer $L$ of size at least $\frac{2^n - 1}{2n}$. Note that below each vertex of $L$ there is a distinct leaf, so $T$ has at least $\frac{2^n - 1}{2n}$ leaves. Since in $G$ every vertex has degree at least two, there must exist at least $\left\lfloor \frac{2^n - 1}{2n} \right\rfloor$ edges that are in $G$ but not in $T$. Hence

$$e(G) \geq e(T) + \frac{2^n - 1}{2n} = 2^n + \frac{2^n - 1}{2n} - 1. \quad \Box$$

A question in the same spirit as Question 9 concerns 2-additive spanners. Denote by $f_2(n)$ the fewest possible edges in a graph $G \subset Q_n$ that is a 2-additive spanner, that is, $d_G(x, y) \leq d_{Q_n}(x, y) + 2$ for all $x, y$. The best known bounds, given in [7, 16] are

$$c_2^n \log n \leq f_2(n) \leq C_2^n \sqrt{n}.$$
Acknowledgements

We would like to express our gratitude to the referee for carefully reading the manuscript and for suggesting several specific improvements.

References


