ELSEVIER

Discrete Mathematics 186 (1998) 289–293

DISCRETE MATHEMATICS

Note

On embedding complete graphs into hypercubes

Michael Klugerman[1], Alexander Russell[2], Ravi Sundaram[3]

Laboratory for Computer Science, MIT, NE43-372, 545 Technology Square, Cambridge, MA 02139, USA

Received 16 January 1996; revised 13 November 1996; accepted 29 September 1997

Abstract

An embedding of K_n into a hypercube is a mapping, ϕ, of the n vertices of K_n to distinct vertices of the hypercube. The associated cost is the sum over all pairs of vertices, v_i, v_j, $i \leqslant j$, of the (Hamming) distance between $\phi(v_i)$ and $\phi(v_j)$. Let $f(n)$ denote the minimum cost over all embeddings of K_n into a hypercube (of any dimension). In this note we prove that $f(n) = (n-1)^2$ unless $n = 4$ or 8, in which case $f(n) = (n-1)^2 - 1$. As an application, we use this theorem to derive an alternate proof of the fact that the Isolation Heuristic (and its accompanying variants) for the multiway cut problem of Dahlhaus et al. (1994) are tight for all n. This result also gives a combinatorial justification for the seemingly anomalous improvements that these variants achieve in the cases $n = 4$ and 8. © 1998 Published by Elsevier Science B.V. All rights reserved

1. Preliminaries

K_n denotes the complete graph on n vertices with unit weight on all edges. The hypercube of dimension n has 2^n vertices, each vertex being labeled with a string of 0's and 1's of length n. The Hamming distance, $\lceil(v_i, v_j)$ between two vertices v_i, v_j, of the hypercube is the number of positions in which the labels of the two vertices are different. The hypercube has edges between every pair of vertices which are at Hamming distance 1. The Hamming distance between a pair of vertices is easily seen to be the length of the shortest path between them. The Hamming weight of a vertex is the number of 1's in its label. A cut of a graph is a subset of vertices of the graph. The edges of a cut are the set of edges which go from vertices within the subset to vertices outside. The weight of a cut is the sum of the weights of the edges of the cut. The weight of a cut collection is simply the sum of the weights of the cuts in the collection.

[1] Escher Group.
[2] Department of Mathematics, MIT.
[3] Delta Global Trading LP.

PII S0012-365X(97)00239-2

2. Result

Theorem 1. *For a set V of n* distinct *vertices $v_1, \ldots, v_n$ in a hypercube (of any dimension) let $f(V) = \sum_{i<j} \lceil (v_i, v_j)$. Then*

$$f(n) = \min_{\{V:\, |V|=n\}} f(V) = \begin{cases} (n-1)^2 - 1 & \text{if } n = 4, 8, \\ (n-1)^2 & \text{otherwise.} \end{cases}$$

Proof. It is clear that the (claimed) minimum in the case of $n=4$ and 8 can be achieved by letting V be the set of all vertices in a two-dimensional and three-dimensional hypercube, respectively. For all other n the (claimed) minimum can be achieved by taking a vertex and its $(n-1)$ neighbours in an $n-1$-dimensional hypercube.

It remains to show that we cannot do better. Consider any set with e nodes of even Hamming weight and o nodes of odd Hamming weight, $e+o=n$. It is clear that we have a lower bound of $(n-1)^2$ if either e or o is 0 or 1. Assume, then, without loss of generality, that $o \geqslant e \geqslant 2$. To prove our bound, we consider the (Hamming) distances between even weight nodes, between odd weight nodes, and then between even and odd weight nodes. Any pair of even weight nodes has distance at least 2. Hence, we have a lower bound of $2\binom{e}{2}$ on the contribution to the sum from pairs of even weight nodes. Similarly, for the odd weight nodes we get a lower bound of $2\binom{o}{2}$. Consider the distances between even and odd weight nodes. For each pair of even weight nodes e_1 and e_2, there are at most two nodes o_1 and o_2 which are each at distance 1 from e_1 and e_2. The remaining $o-2$ odd weight nodes are each an average distance of at least 2 from both e_1 and e_2. Based on these averages we arrive at a lower bound for the minimum total Hamming distance of

$$2\binom{e}{2} + 2\binom{o}{2} + \left(\frac{2(o-2)+2)}{o}\right) eo = n^2 - n - 2e.$$

Note that $n^2 - n - 2e \geqslant n^2 - 2n$. Hence, if $e \neq o$, we arrive at a lower bound of $(n-1)^2$. Now, consider the remaining case when $e = o > 1$. When $e = 2, 4$ we have explicit constructions which match the lower bound of $n^2 - 2n$. It remains to consider the cases $e = o = 3$ and $e = o \geqslant 5$.

In order for the above (lower-bound) argument concerning inter-node distances to be tight, the set must have special structure. Specifically, any pair of nodes having the same parity must have distance 2. This implies that all the nodes with the same parity lie in a Hamming ball of radius 1 (in fact, this holds for all $e = 0 \neq 4$). Hence, without loss of generality, one may assume that all odd nodes have weight exactly 1. Since $e \geqslant 3$ there are at least two even nodes with weight 2 or more. Clearly, there can be at most one odd node that is at distance 1 from both of them, instead of two nodes as counted in the above lower bound argument. Hence, the lower bound is not tight. $\square$

3. Related results

Our main result has been investigated previously under the subject "average distance". We refer the interested reader to [1] and the references therein for an extensive literature on this subject. Althofer and Sillke [1] prove a related result, but one that is not as sharp as the one obtained in this paper.

4. Application

Our original motivation for solving the problem of embedding complete graphs in hypercubes arose from the *multiway* or *n-way cut* problem. In the n-way cut problem we are given an edge-weighted graph and n distinguished vertices called terminals; we are required to find a minimum weight n-way cut, i.e. a set of edges whose removal separates every terminal pair. This problem is simply the min-cut max-flow problem when $n=2$. In [2] it was shown that the problem becomes NP-hard for $n=3$. They gave a simple approximation algorithm, the *isolation heuristic*, for arbitrary graphs that comes within a factor of $2(1-(1/n))$ of the optimal. They also gave variants of the isolation heuristic which do better for $n=4$ and 8. They state in the paper, without proof, that similar approaches are bound to fail for all other values of n.

For the sake of completeness we present the isolation heuristic and a proof of its performance guarantee. We also present its variants which achieve an improved factor for $n=4$ and 8.

Isolation heuristic: 1. For $1 \leqslant i \leqslant n$ construct a minimum weight isolating cut E_i for terminal s_i, i.e. a min-cut that separates s_i from the other terminals.

2. Let E be the union of the cheapest $n-1$ of the cuts E_i. Return E.

Lemma 1 (Dahlhaus et al. [2]). *The isolation heuristic constructs an n-way cut whose weight is guaranteed to be no more than* $2-(2/n)$ *times the optimal.*

Proof. Let $\hat{E}$ be an optimal n-way cut. Let $w(\hat{E})$ denote the sum of the weights of the edges in the cut. For $1 \leqslant i \leqslant n$, let $\hat{V}_i$ be the set of vertices left connected to s_i by $\hat{E}$ and let $\hat{E}_i$ be the set of edges in $\hat{E}$ with one endpoint in $\hat{V}_i$. Observe that for each i the set $\hat{E}_i$ is an isolating cut for s_i. Thus, $w(\hat{E}_i) \geqslant w(E_i)$. Since, $\sum_{i=1}^{n} w(\hat{E}_i) = 2w(\hat{E})$ we have that

$$w(E) \leqslant \frac{n-1}{n} \sum_{i=1}^{n} w(E_i) \leqslant \frac{n-1}{n} \sum_{i=1}^{n} w(\hat{E}_i) \leqslant \left(2-\frac{2}{n}\right) w(\hat{E}). \qquad \square$$

Variant for the case $n=4$: For $n=4$ the isolation heuristic given above provides a performance guarantee of $2-(2/n) = 2-\frac{2}{4} = \frac{3}{2}$. An improved guarantee of $\frac{4}{3}$ can be obtained by the following: for each partition of the terminals into sets S_1, S_2 of size two, use max-flow techniques to compute the minimum cut that separates the terminals in S_1 from those in S_2; output the union of (the best) two such cuts. It is an easy matter to prove that this scheme achieves a performance guarantee of $\frac{4}{3}$.

Variant for the case $n=8$: For $n=8$ the isolation heuristic given above provides a performance guarantee of $2-(2/n)=2-\frac{2}{8}=\frac{7}{4}$. An improved guarantee of $\frac{12}{7}$ can be obtained by the following: for each partition of the terminals into sets S_1, S_2 of size four, use max flow techniques to compute the minimum cut that separates the terminals in S_1 from those in S_2; there exists a set of three of these cuts whose union is an 8-way cut and whose total weight is no more than the average; output the union of (the best) three such cuts. It is an easy matter to prove that this scheme achieves a performance guarantee of $\frac{12}{7}$.

The isolation heuristic (and its variants for the cases $n=4$ and 8) can be thought of as essentially finding a minimum weight collection of cuts that separates all pairs of terminals. In what follows, we show that the isolation heuristic is tight. There is a combinatorial basis for the seemingly anomalous improvements in the cases $n=4$ and 8 and these cannot be extended to any other n.

Lemma 2. *If K_n has a cut collection of weight C separating all pairs of vertices then there exists a cut collection for any instance of the n-way cut problem separating all pairs of terminals with weight at most $2C/n(n-1)$ times the weight of the optimal n-way cut.*

Proof. The proof is a straightforward averaging argument.

Let $\hat{E}$ be an optimal n-way cut. Let $w(\hat{E})$ denote the sum of the weights of the edges in the cut. For $1 \leqslant i \leqslant n$, let $\hat{V}_i$ be the set of vertices left connected to s_i by $\hat{E}$ and let $\hat{E}_{ij}$ be the set of edges in $\hat{E}$ with one endpoint in $\hat{V}_i$ and one in $\hat{V}_j$. Observe that for any set S of terminals in the n-way cut problem graph the weight of the minimum cut separating S from the complement set of terminals is at most $\sum_{i\in S,\, j\notin S} w(\hat{E}_{ij})$.

We associate each vertex of K_n one-to-one to a terminal of the graph in the n-way cut problem. Any cut of K_n separates a subset of the vertices from its complement. A particular cut of K_n we map to the min-weight cut of the n-way cut problem graph that separates the corresponding set of terminals from the complement set of terminals. In this way we have a mapping from cuts, (and hence cut collections) of K_n to cuts, (and cut collections) of the n-way cut problem graph. Consider all possible mappings of the vertices of K_n one to one to terminals of the graph. There are $n!$ such mappings. For each mapping consider the weight of the collection of min-weight cuts of the n-way cut problem corresponding to the cut collection of weight C in K_n. The sum over all $n!$ mappings of the weight of these cut collections is at most $n!C\sum_{ij} w(\hat{E}_{ij})/\binom{n}{2}$.

Since the average weight is within a factor of $2C/n(n-1)$ of the optimal for the n-way cut problem there exists a mapping, and hence a cut collection, which achieves this bound. □

By the above lemma we see that the best performance guarantee achievable by heuristics (like the isolation heuristic) for the n-way cut problem that output cut collections is dependent directly on the minimum weight of a cut collection in K_n that separates all pairs of vertices.

Lemma 3. *The minimum weight of any cut collection that separates all pairs of vertices of K_n is equal to $f(n)$.*

Proof. Given any cut collection $\mathscr{C} = \{c_1, c_2, \ldots, c_k\}$ we can create an embedding of equivalent weight in a hypercube of dimension k. We have one dimension per cut and a vertex of K_n gets mapped to that vertex of the hypercube with a 1 in the ith position of the label iff the original vertex of K_n is in the ith cut. It is easy to see that if V is the set of mapped vertices then $f(V)$ is equal to the weight of $\mathscr{C}$.

Similarly, given any embedding V in a hypercube of dimension k one can create a cut collection of equivalent weight by having one cut for each dimension and putting all those mapped vertices in the cut which have a 1 in the label at the dimension corresponding to the cut. □

Corollary 1. *The isolation heuristic and its variants are tight. Consider heuristics which output cut collections as solutions to the n-way cut problem. If in the analysis of the performance guarantee the weight of the cut collection is measured against that of the optimal, then the best that such heuristics can achieve is a factor of* $2(1-(1/n))$, *except when* $n=4$ *or* 8 *in which case they can achieve a solution that is at most* $2(1-(1/n-1))$ *times the optimal (to the n-way cut problem).*

Proof. Follows from Theorem 1 and Lemmas 2 and 3. □

Acknowledgements

We thank Michel Goemans for suggesting the problem and for useful discussions.

References

[1] I. Althofer, T. Sillke, An 'average distance' inequality for large subsets of the cube, J. Combin. Theory Ser. B 56 (1992) 296–301.

[2] E. Dahlhaus, D.S. Johnson, C.H. Papadimitriou, P.D. Seymour, M. Yannakakis, The complexity of multiway cuts, Proc. 24th Ann. ACM Symp. on the Theory of Computing, 1992, pp. 241–251. Journal version The complexity of multiterminal cuts, SIAM J. Comput. 23 (4) (1994) 864–894.