On bipartite coverings of graphs and multigraphs

Noga Alon *

Abstract

A bipartite covering of a (multi)graph G is a collection of bipartite graphs, so that each edge of G belongs to at least one of them. The capacity of the covering is the sum of the numbers of vertices of these bipartite graphs. In this note we establish a (modest) strengthening of old results of Hansel and of Katona and Szemerédi, by showing that the capacity of any bipartite covering of a graph on n vertices in which the maximum size of an independent set containing vertex number i is α_i, is at least $\sum_i \log_2(n/\alpha_i)$. We also obtain slightly improved bounds for a recent result of Kim and Lee about the minimum possible capacity of a biparite covering of complete multigraphs.

1 Introduction

A bipartite covering $\mathcal{H} = \{H_1, \dots, H_m\}$ of a graph G on the set of vertices $[n] = \{1, 2, \dots, n\}$ is a collection of bipartite graphs H_i on $[n]$, so that each edge of G belongs to at least one of them. Note that each H_i is not necessarily a subgraph of G, the only assumption is that it is a bipartite subgraph of the complete graph on $[n]$. The capacity $\text{cap}(\mathcal{H})$ of the cover is the sum $\sum_i |V(H_i)|$ of the numbers of vertices of these bipartite graphs. A known result of Hansel [1] is that the capacity of any bipartite covering of the complete graph K_n on n vertices is at least $n \log_2 n$. This bound is tight when n is power of 2.

In [2] Kim and Lee consider the analogous problem, where the complete graph K_n is replaced by the complete multigraph K_n^λ in which every pair of distinct vertices is connected by λ parallel edges. A bipartite covering here is a collection of bipartite graphs so that each edge belongs to at least λ of them. They prove that the capacity of each bipartite covering of K_n^λ is at least

$$\max\{2\lambda(n-1), n[\log n + \lfloor(\lambda-1)/2\rfloor \log(\frac{\log n}{\lambda}) - \lambda - 1]\},$$

*Princeton University, Princeton, NJ 08544, USA and Tel Aviv University, Tel Aviv 69978, Israel. Email: nalon@math.princeton.edu. Research supported in part by NSF grant DMS-2154082.

where all logarithms here and in the rest of this note are in base 2. They also establish an upper bound: there exists a bipartite covering of K_n^λ of capacity at most

$$n(\log(n-1) + (1+o(1))\lambda \log\log n).$$

This shows that for $\lambda = (\log n)^{1-\Omega(1)}$ the smallest possible capacity is $n\log n + \Theta(n\lambda \log\log n)$ but leaves an additive gap of $\Omega(\lambda n \log\log n)$ between the upper and lower bounds.

The proofs in [2] proceed by studying a more general problem regarding graphons. Our first contribution in this note is a shorter combinatorial proof of the results above, slightly improving the bounds. Let $\mathrm{cap}(n,\lambda)$ denote the minimum possible capacity of a bipartite covering of K_n^λ.

Theorem 1.1 (Lower bound)**.** *For positive integers $n \geq 2$ and λ,*

$$cap(n,\lambda) \geq \max\{2\lambda(n-1), n[\log n + \lfloor(\lambda-1)/2\rfloor \log(\frac{2\log n}{\lambda-1})]\}$$

Theorem 1.2 (Upper bound)**.** *Let $k(n,\lambda)$ denote the minimum length of a binary error correcting code with distance at least λ which has at least n codewords. Then $cap(n,\lambda) \leq n \cdot k(n,\lambda)$. Therefore*

1. *For any $n \geq 2$*

$$cap(n,2) \leq n(\lceil \log n \rceil + 1) < n(\log n + 2).$$

2. *For any n and $\lambda \leq 0.5\log n$*

$$cap(n,\lambda) \leq n[\log n + (\lambda-1)(\log(\frac{\log n}{\lambda-1}) + 4)]$$

3. *For any $0 < c < 1/2$, and for $\lambda \geq c\frac{\log n}{1-H(c)}$ where $H(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function,*

$$cap(n,\lambda) \leq \frac{\lambda}{c} n.$$

4. *For any fixed λ there are infinitely many values of n so that*

$$cap(n,\lambda) \leq n[\log n + \lfloor(\lambda-1)/2)\rfloor \log\log n + 2].$$

Katona and Szemerédi [3] proved the following generalization of the result of Hansel, dealing with the capacity of bipartite coverings of general graphs.

Theorem 1.3 ([3]). *Let G be a graph on the set of vertices $[n]$ and let $d_1, d_2, \ldots, d_n$ denote the degrees of its vertices. Then the capacity of any bipartite covering of G is at least*

$$\sum_{i=1}^{n} \log(\frac{n}{n-d_i}).$$

Our second contribution here is the following strengthening of this result.

Theorem 1.4. *Let G be a graph on the set of vertices $[n]$. For each vertex i let α_i denote the maximum size of an independent set of G that contains the vertex i. Then the capacity of any bipartite covering of G is at least*

$$\sum_{i=1}^{n} \log(\frac{n}{\alpha_i}).$$

Since it is clear that $\alpha_i \leq n - d_i$ for every i, this is indeed a strengthening of the Katona-Szemerédi result (Theorem 1.3). The binomial random graph $G = G(n, 0.5)$ is one example for which Theorem 1.4 is strictly stronger than Theorem 1.3. Indeed, with high probability for $G = G(n, 0.5)$, $d_i = (1/2 + o(1))n$ for every i and $\alpha_i = (2 + o(1)) \log n$ for every i. Therefore the lower bound of Theorem 1.3 for this G is typically $(1 + o(1))n$, whereas the lower bound provided by Theorem 1.4 is $n \log n - (1 + o(1))n \log \log n$. This is tight since the chromatic number of $G = G(n, 0.5)$ is, with high probability, $\chi_n = (1 + o(1))\frac{n}{2 \log n}$ implying that G admits a bipartite covering consisting of $\lceil \log \chi_n \rceil$ spanning bipartite graphs, and the corresponding capacity is $n \log n - (1 + o(1))n \log \log n$.

The rest of this note contains the (short) proofs of the results above.

2 Complete multigraphs

2.1 The lower bound

Let $\mathcal{H} = \{H_1, \ldots, H_m\}$ be a bipartite covering of the complete multigraph K_n^λ on the set of vertices $[n] = \{1, 2, \ldots, n\}$. We first prove that $\text{cap}(\mathcal{H}) \geq 2\lambda(n-1)$. Let n_i denote the number of vertices of H_i. Since it is bipartite the number of its edges is at most $n_i^2/4$. As the edges of all these graphs cover each of the $n(n-1)/2$ edges of the complete graph on $[n]$ at least λ times, and since $n_i \leq n$ for all i, it follows that

$$\frac{n}{4} \sum_{i=1}^{m} n_i \geq \sum_{i=1}^{m} \frac{n_i^2}{4} \geq \lambda n(n-1)/2.$$

This implies that $\text{cap}(\mathcal{H}) = \sum_{i=1}^{m} n_i \geq 2\lambda(n-1)$, as needed.

Note that for any even n this inequality is tight for infinitely many (large) values of λ. In particular it is tight for $\lambda = \frac{n}{4n-4}\binom{n}{n/2}$, and if there is a Hadamard matrix of order n then it is tight for $\lambda = n/2$ as well. In addition, if for some fixed n it is tight for λ_1 and λ_2 then it is also tight for their sum $\lambda = \lambda_1 + \lambda_2$.

We next prove the second inequality, that

$$\text{cap}(\mathcal{H}) \geq n[\log n + \lfloor(\lambda-1)/2\rfloor \log(\frac{2\log n}{\lambda - 1})].$$

Without loss of generality assume that each of the bipartite graphs H_i in $\mathcal{H}$ is a complete bipartite graph, and let $L_i, R_i \subset [n]$ denote its two color classes. For each vertex $j \in [n]$ let A_j denote the set of indices i for which the vertex j belongs to the vertex class L_i of H_i and let B_j be the set of indices i for which $j \in R_i$. Let $x_j = |A_j| + |B_j|$ be the total number of bipartite graphs H_i that contain the vertex j. Note that $x_j \geq \lambda$ for each j, as any edge incident with j must be covered at least λ times.

Put $r = \lfloor(\lambda-1)/2\rfloor$ and let $v = (v_1, v_2, \ldots, v_m\}$ be a uniform random binary vector of length m. For each j, $1 \leq j \leq n$, let E_j denote the event that the number of indices i that belong to A_j for which $v_i = 1$ plus the number of indices i that belong to B_j for which $v_i = 0$ is at most r. It is clear that the probability of E_j is exactly the probability that the binomial random variable $B(x_j, 1/2)$ is at most r, which is

$$p(x_j, 1/2) = \frac{\sum_{q=0}^{r}\binom{x_j}{q}}{2^{x_j}}.$$

Note, crucially, that the events E_j are pairwise disjoint. This is because for every two distinct vertices j and j' there are at least $\lambda > 2r$ indices i for which j and j' belong to the two distinct vertex classes of H_i. Therefore

$$\sum_{j=1}^{n} p(x_j, 1/2) \leq 1.$$

The desired lower bound for $\text{cap}(\mathcal{H}) = \sum_{j=1}^{n} x_j$ can be deduced from the last inequality by a convexity argument. We proceed with the details. Note, first, that for every x, $p(x, r) \geq (x/r)^r 2^{-x}$. Therefore

$$\sum_{j=1}^{n} (x_j/r)^r 2^{-x_j} \leq 1. \tag{1}$$

Recall also that for each j, $x_j \geq \lambda \geq 2r+1$. Consider the function $f(x) = (x/r)^r 2^{-x}$. A simple computation shows that for $r = 1$ its second derivative is $(\ln 2)2^{-x}[x \ln 2 - 2]$ which is positive for all $x \geq 2r+1 = 3$. For $r \geq 2$ the second derivative of $f(x)$ is

$$(x/r)^{r-2} 2^{-x}[((x/r)\ln 2 - 1)^2 - 1/r].$$

It is not difficult to check that this is positive for all $x \geq 2r+1$. This shows that $f(x)$ is convex in the relevant range. Therefore, by (1) together with Jensen's Inequality, if we denote $x = \text{cap}(\mathcal{H}) = \sum x_j$ we get

$$n(\frac{x}{nr})^r 2^{-x/n} \leq 1$$

implying that

$$x \geq n[\log n + r\log(\frac{x}{nr})].$$

Since $x/n \geq \log n$ this shows that

$$x \geq n[\log n + r\log(\frac{\log n}{r}] \geq n[\log n + \lfloor(\lambda-1)/2\rfloor \log(\frac{2\log n}{\lambda-1})].$$

This completes the proof of Theorem 1.1. □

2.2 The upper bound

In this subsection we prove Theorem 1.2. Put $k = k(n,\lambda)$ and let $A = (a_{ij})$ be the k by n binary matrix whose columns are n of the codewords of a binary code of length k with minimum distance (at least) λ. For each i, $1 \leq i \leq k$, let H_i be the complete bipartite graph on the classes of vertices $L_i = \{j : a_{ij} = 0\}$ and $R_i = \{j : a_{ij} = 1\}$. It is easy to see that these bipartite graphs cover every edge of the complete graph on $[n]$ at least λ times. The capacity of this covering is at most kn, establishing the first part of the theorem. The subsequent items in the theorem follow by considering appropriate known error correcting codes, see, e.g. [4].

For the first item simply take the code consisting of all 2^{k-1} codewords with even Hamming weight. Since $2^{k-1} \geq n$ for $k = \lceil \log n \rceil + 1$ the claimed result follows. The second and third items follow from the Gilbert-Varshamov bound which gives that the maximum cardinality of a binary code with length k and distance λ is at least

$$\frac{2^k}{\sum_{i=0}^{\lambda-1}\binom{k}{i}}.$$

This quantity is at least $2^k(\frac{ek}{\lambda-1})^{-(\lambda-1)}$, implying the second item. For any $\lambda = ck \leq k/2$ this quantity is also at least $2^{(1-H(c)k}$, where $H(x)$ is the binary entropy function. This yields the third item.

The fourth follows by considering an appropriate augmented BCH code. For any k which is a power of 2 and for any d this is a (linear) binary code of length k with

$$n = \frac{2^k}{2k^{d-1}}$$

codewords and minimum distance $2d$. For $d-1 = \lfloor(\lambda-1)/2\rfloor$, $2d \geq \lambda$ and

$$k = \log n + 1 + (d-1)\log k \leq \log n + \lfloor(\lambda-1)/2\rfloor \log\log n + 2.$$

This completes the proof of Theorem 1.2. □

3 General graphs

In this section we prove Theorem 1.4. We need the following simple lemma.

Lemma 3.1. *Let $E_i, i \in I$ be a finite collection of events in a (discrete) probability space. Suppose that for every point x in the space, if $x \in E_i$ then the total number of events E_j in the collection that contain x is at most a_i. Then*

$$\sum_{i \in I} \frac{Prob(E_i)}{a_i} \leq 1. \tag{2}$$

It is worth noting that the above holds (with the same proof) for any probability space, the assumption that it is discrete here is merely because this is the case we need, and it slightly simplifies the notation in the proof.

Proof. Let x be an arbitrary point of the space, and let $p(x)$ denote its probability. Suppose it belongs to r of the events E_i, let these be $E_{i_1}, \ldots E_{i_r}$. By the definition of the numbers a_i it follows that $a_{i_j} \geq r$ for all $1 \leq j \leq r$. Therefore the total contribution of the point x to the sum in the left-hand-side of (2) is

$$\sum_{j=1}^{r} \frac{p(x)}{a_{i_j}} \leq \sum_{j=1}^{r} \frac{p(x)}{r} \leq p(x).$$

The desired result follows by summing over all points x in the space. □

Proof of Theorem 1.4: Let G be a graph on the set of vertices $[n]$, let α_i denote the maximum cardinality of an independent set of G containing the vertex i, and let $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ be a bipartite covering of G.

As in the proof of Theorem 1.1 we may and will assume, without loss of generality, that each of the bipartite graphs H_i in $\mathcal{H}$ is a complete bipartite graph. Let $L_i, R_i \subset [n]$ denote its two color classes. For each vertex $j \in [n]$ let A_j denote the set of indices i for which the vertex j belongs to the vertex class L_i of H_i and let B_j be the set of indices i for which $j \in R_i$. Let $x_j = |A_j| + |B_j|$ be the total number of bipartite graphs H_i that contain the vertex j. Our objective is to prove a lower bound for the capacity of $\mathcal{H}$, which is exactly the sum $\sum_{j=1}^{n} x_j$.

Let $v = (v_1, v_2, \ldots, v_m\}$ be a uniform random binary vector of length m. For each j, $1 \leq j \leq n$, let E_j denote the event that $v_i = 0$ for every index i that belongs to A_j and $v_i = 1$ for every index i that belongs to B_j. Note that the probability of E_j is exactly 2^{-x_j}. Note also that if some point $v = (v_1, v_2, \ldots, v_m\}$ belongs to the events $E_j, j \in J$, then the set of vertices $J \subset [n]$ is an independent set of G. Indeed, if some two vertices in J are adjacent, then the edge connecting them belongs to at least one of the graphs H_i implying that one of these vertices belongs to L_i whereas the other lies in R_i and showing that they can't both satisfy the requirement given by v_i. It thus follows that any point v that lies in E_j belongs to at most α_j of the events $E_{j'}$. Therefore, by Lemma 3.1

$$\sum_{j=1}^{n} 2^{-x_j - \log \alpha_j} = \sum_{j=1}^{n} \frac{2^{-x_j}}{\alpha_j} \leq 1.$$

By the arithmetic-geometric means inequality this implies

$$n 2^{-(\sum_{j=1}^{n} x_j + \sum_{j=1}^{n} \log \alpha_j)/n} \leq 1,$$

giving

$$2^{\sum_{j=1}^{n} x_j} \geq n^n 2^{-\sum_{j=1}^{n} \log \alpha_j} = 2^{\sum_{j=1}^{n} (\log n - \log \alpha_j)}.$$

Therefore

$$\sum_{j=1}^{n} x_j \geq \sum_{j=1}^{n} \log(\frac{n}{\alpha_j}),$$

completing the proof. □

References

[1] Georges Hansel, Nombre minimal de contacts de fermeture nécessaires pour réaliser une fonction booléenne symétrique de n variables, C. R. Acad. Sci. Paris 258 (1964), 6037–6040.

[2] Jaehoon Kim and Hyunwoo Lee, Covering multigraphs with bipartite graphs, arXiv:2304.11691, 2023.

[3] Gyula Katona and Endre Szemerédi, On a problem of graph theory, Studia Sci. Math. Hungar. 2 (1967), 23–28.

[4] F. Jessie MacWilliams and Neil J. A. Sloane, *The Theory of Error-Correcting Codes*, I. North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford (1977).