Studies in Pure Mathematics
To the Memory of Paul Turán

On the decomposition of graphs into complete bipartite subgraphs

by

F. R. K. CHUNG (Murray Hill), P. ERDŐS (Budapest)
and J. SPENCER* (Stony Brook)

Abstract

For a given graph G, we consider a **B**-decomposition of G, i.e., a decomposition of G into complete bipartite subgraphs $G_1, \ldots, G_t$, such that any edge of G is in exactly one of the G_i's. Let $\alpha(G; \mathbf{B})$ denote the minimum value of $\sum_i |V(G_i)|$ over all **B**-decompositions of G. Let $\alpha(n; \mathbf{B})$ denote the maximum value of $\alpha(G; \mathbf{B})$ over all graphs on n vertices.

A **B**-covering of G is a collection of complete bipartite subgraphs $G'_1, G'_2, \ldots, G'_t$, such that any edge of G is in at least one of the G'_i. Let $\beta(G; \mathbf{B})$ denote the minimum value of $\sum_i |V(G'_i)|$ over all **B**-coverings of G and let $\beta(n; \mathbf{B})$ denote the maximum value of $\beta(G; \mathbf{B})$ over all graphs on n vertices.

In this paper, we show that for any positive ε, we have

$$(1-\varepsilon)\frac{n^2}{2e\log n} < \beta(n; \mathbf{B}) \leqq \alpha(n; \mathbf{B}) < (1+\varepsilon)\frac{n^2}{2\log n}$$

where $e = 2.718\ldots$ is the base of natural logarithms, provided n is sufficiently large.

Introduction

For a finite graph G, a *decomposition* P of G is a family of subgraphs $G_1, G_2, \ldots, G_t$, such that any edge in G is an edge of exactly one of the G'_i s. If all G'_i s belong to a specified class of graphs **H**, such a decomposition will be called an **H**-decomposition of G (see [2]).

Let f denote a *cost function* for graphs which assigns certain non-negative real values to all graphs. Sometimes it is desirable to decompose a given graph into subgraphs in **H** such that the total "cost" (the sum of the cost function values of all subgraphs) is minimized. In other words, for a given graph G, we consider the following:

* Work done while a consultant at Bell Laboratories.

$$\alpha_f(G;\mathbf{H}) = \min_P \sum_i f(G_i)$$

where $P = \{G_1, G_2, \ldots, G_t\}$ ranges over all **H**-decompositions of G.

Also of interest to us will be the quantity

$$\alpha_f(n;\mathbf{H}) = \max_G \alpha_f(G;\mathbf{H})$$

where G ranges over all graphs on n vertices.

If we take f_0 to be the counting function, which assigns value 1 to any graph, and **P** is the family of all planar graphs, then $\alpha_{f_0}(G; \mathbf{P})$ is simply the thickness of G. If **F** denotes the family of forests, then $\alpha_{f_0}(G; \mathbf{F})$ is called the arboricity of G (see [6]). Many results along these lines are available. The reader is referred to [2] for a brief survey.

In this paper, we will deal almost exclusively with the case in which **H** is **B**, the family of complete bipartite graphs. By a theorem in [5], the value of $\alpha_{f_0}(n; \mathbf{B})$ is given by:

$$\alpha_{f_0}(n;\mathbf{B}) = n-1\,.$$

We consider the cost function f_1 where the value $f_1(G)$ is just the number of vertices in G. In the remaining part of the paper, we abbreviate $\alpha(n) = \alpha_{f_1}(n; \mathbf{B})$ and $\alpha(G) = \alpha_{f_1}(G; \mathbf{B})$. In particular, we show for any given ε and sufficiently large n,

$$(1) \qquad (1-\varepsilon)\frac{n^2}{2e\log n} < \alpha(n) < (1+\varepsilon)\frac{n^2}{2\log n}$$

where e satisfies $\ln e = 1$.

An **H**-covering of G is a collection of subgraphs of G, say $G'_1, \ldots, G'_{t'}$, such that any edge of G is in at least one of the G'_i, and all G'_i are in **H**. For a given cost function f, we can define

$$\beta_f(G;\mathbf{H}) = \min_P \sum_i f(G'_i)$$

where $P = \{G'_1, \ldots, G'_{t'}\}$ ranges over all **H**-coverings of G.

It is easily seen that

$$\beta_f(G;\mathbf{H}) \leqq \alpha_f(G;\mathbf{H})$$

and

$$\beta_f(n;\mathbf{H}) \leqq \alpha_f(n;\mathbf{H})\,.$$

We will show that the asymptotic growth of $\beta_{f_1}(n; \mathbf{B})$ is quite similar to $\alpha_{f_1}(n; \mathbf{B})$. In fact, we will obtain the same upper and lower bounds for $\beta_{f_1}(n; \mathbf{B})$ as those for $\alpha_{f_1}(n; \mathbf{B})$ in (1).

A lower bound

We derive these bounds mainly by probabilistic methods, which have been extensively described in the book by two of the authors [4].

Theorem 1. $\alpha(n) \geqq (1-\varepsilon)\dfrac{n^2}{2e\log n}$ *for any given positive* ε *and sufficiently large* n.

Proof. Let us consider a random graph G with n vertices and $\lfloor n^2/2e \rfloor$ edges. The probability of G containing a complete bipartite subgraph $K_{a,b}$ is bounded above by

$$\binom{n}{a}\binom{n}{b} e^{-ab} < e^{(a+b)\log n - ab}$$

(where $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the greatest integer less than x and the least integer greater than x, respectively.)

Let S denote the set of all unordered pairs $\{a, b\}$ satisfying

$$1 \leqq a, b \leqq n, \frac{a+b}{ab} < \frac{1-\varepsilon}{\log n}.$$

The probability of G containing one of the complete bipartite subgraphs $K_{a,b}$ with $\dfrac{a+b}{ab} < \dfrac{1-\varepsilon}{\log n}$ is bounded above by

$$\sum_{\{a,b\}\in S} \binom{n}{a}\binom{n}{b} e^{-ab} < \sum_{\{a,b\}\in S} e^{-\varepsilon ab} < \sum_{\{a,b\}\in S} e^{-\varepsilon(\log n)^2} < n^2 e^{-\varepsilon(\log n)^2} < 1$$

for large n.

Therefore, there exists a graph G with n vertices and $\lfloor n^2/2e \rfloor$ edges such that G does not contain any $K_{a,b}$ as a subgraph. Let $P = \{G_1, G_2, \ldots, G_t\}$ denote a $\mathbf{B}$-decomposition of G such that $\alpha(G)$ is the sum of the sizes of vertex set $V(G_i)$ of G_i, i.e.,

$$\alpha(G) = \sum_{i=1}^{t} |V(G_i)|.$$

For any edge (u, v) in G, we define

$$f(u, v) = \frac{|VG_i)|}{|E(G_i)|}$$

where $\{u, v\}$ is in $E(G_i)$, the edge set of G_i.

It is easily seen that

$$\alpha(G) = \sum_{\{u,v\}} f(u, v).$$

Since G does not contain $K_{a,b}$ as a subgraph, any $G_i = K_{c,d}$, $1 \leqq i \leqq t$, satisfies that $\frac{c+d}{cd} \geqq \frac{1-\varepsilon}{\log n}$. Thus we have

$$f(u,v) \geqq \frac{1-\varepsilon}{\log n} \quad \text{for } any\ \{u,v\}\ in\ E(G).$$

and

$$\alpha(n) > \alpha(G) > \frac{(1-\varepsilon)n^2}{2e \log n}$$

for sufficiently large n. This proves the theorem.

An upper bound

First, we shall prove a preliminary result.

Lemma. *For any $\varepsilon > 0$ any graph on n vertices and $\rho\binom{n}{2}$ edges contains a complete bipartite graph $K_{s,t}$ as a subgraph where $t = \lfloor 1(1-\varepsilon)n\rho^s \rfloor$ and $s < \varepsilon\rho n$ for n sufficiently large.*

Proof. Suppose G has n vertices and $\rho\binom{n}{2}$ edges and G does not contain $K_{s,t}$ as a subgraph. From the proof in [3], the following holds:

$$n(\rho n - s)^s \leqq (t-1)\cdot n^s. \tag{2}$$

However, on the other hand, we have

$$(t-1)n^s < tn^s \leqq (1-\varepsilon)n^{1+s}\rho^s < n(\rho n - s)^s$$

since $s < \varepsilon\rho n$.

This contradicts (2). Thus G must contain $K_{s,t}$.

Theorem 2. *For any given ε, we have*

$$\alpha(n) < (1+\varepsilon)\frac{n^2}{2\log n} \tag{3}$$

if n is large enough.

Proof. From Lemma 1, one can easily verify that a graph G on $\rho\binom{n}{2}$ edges and n vertices contains a subgraph H isomorphic to $K_{s,t}$, where $s = \lfloor (1-\varepsilon_1) \log n \log(1/\rho) \rfloor$ and $t = \lfloor s^2 \log(1/\rho) \rfloor$ and $\varepsilon_1 > \frac{(\log n)^2}{\rho n}$. We will decompose G into complete bipartite subgraphs by a "greedy algorithm". Given G we find a subgraph H isomorphic to $K_{s,t}$ and let G_1 to be the subgraph of G containing all edges of G except those in H. Now, we find a subgraph H_1 isomorphic to K_{s_1,t_1} and let G_2 to be a subgraph of G_1 containing all edges of G_1 except those in H_1 and continue in this fashion until only $\varepsilon_2 \frac{n^2}{\log n}$ edges are left. Thus G is decomposed into $H, H_1, \ldots$, together with $\varepsilon_2 \frac{n^2}{\log n}$ edges and we have the following recursive relation

(4) $$\alpha(G) \leqq s + t + \alpha(G_1).$$

We will prove by induction that for a give $\varepsilon < \varepsilon_2 < \varepsilon_1$, $\varepsilon_3 > 0$ and sufficiently large n the following holds,

(5) $$\alpha(G) \leqq (1+\varepsilon_2) \frac{n^2}{2 \log n} \int_0^\rho \log(1/x)\, dx + 2\varepsilon_2 \frac{n^2}{\log n}.$$

Suppose (5) holds for any graph H with $|E(H)| < \rho \binom{n}{2}$. From (4), we have

$$\alpha(G) \leqq (1-\varepsilon_2)(\log n)^2/(\log(1/\rho))^3 + (1+\varepsilon_2) \frac{n^2}{2 \log n} \int_0^{\rho'} \log(1/x) dx + 2\varepsilon_2 \frac{n^2}{\log n}$$

where $\rho' = (|E(G)| - st) \Big/ \binom{n}{2}$ for n sufficiently large. It suffices to show that

$$(1-\varepsilon_2)\ (\log n)^2 / \log(1/\rho))^3 + (1+\varepsilon_2) \frac{n^2}{2 \log n} \int_0^{\rho'} \log(1/x)\, dx \leqq$$

$$\leqq (1+\varepsilon_2) \frac{n^2}{2 \log n} \int_0^{\rho'} \log(1/x)\, dx$$

This can be verified by straightforward calculation. Thus (5) is proved and we have

$$\alpha(n) \leqq (1+\varepsilon_2) \frac{n^2}{2 \log n} \int_0^1 \log(1/x) dx + 2\varepsilon_2 \frac{n^2}{\log n} \leqq (1+\varepsilon) \frac{n^2}{2 \log n}$$

for given $\varepsilon > 0$. Theorem 2 is proved.

By slightly modifying the proofs of Theorem 1, we can easily prove the following.

Theorem 3.

$$\beta_{f_1}(n; \mathbf{B}) \geqq (1-\varepsilon) \frac{n^2}{2e \log n}$$

for any positive ε *and sufficiently large* n.

Therefore we have

$$(1-\varepsilon)\frac{n^2}{2e\log n}<\beta_{f_1}(n;\mathbf{B})\leqq\alpha_{f_1}(n;\mathbf{B})<(1+\varepsilon)\frac{n^2}{2\log n}$$

for any given positive ε and sufficiently large n, which summarizes the main results of the paper.

Some related question

As we noted earlier, the lower bound is obtained by a probabilistic method which is nonconstructive. It would be of great interest to find an explicit construction of a graph G on n vertices, $c_1n^2/\log n$ edges (or c_2n^2 edges) which does not contain an $K_{c_3\log n, c_3\log n}$ as a subgraph for some constants c_1, c_2 and c_3.

Another interesting problem which has long been conjectured [4] concerns the Turán number $T(K_{t,t};n)$, the maximum number of edges a graph on n vertices can have which does not contain $K_{t,t}$ as a subgraph. Is it true that

$$T(K_{t,t};n)=O(n^{2-1/t})?$$

For the case $t=3$, the above equality has been verified in [1].

In this paper, we have shown that $\alpha_{f_1}(n;\mathbf{B})=O(n^2/\log n)$. However, we do not know the existence of

$$\lim_{n\to\infty}\frac{\alpha_{f_1}(n;\mathbf{B})}{n^2/\log n}\quad\text{or}\quad\lim_{n\to\infty}\frac{\beta_{f_1}(n;\mathbf{B})}{n^2/\log n},$$

obviously.

Let $\mathbf{G}_n$ be the set of all the $2^{\binom{n}{2}}$ labelled graphs on n vertices. It would be of interest to evaluate $\sum_{G\in\mathbf{G}_n}\alpha_{f_1}(G;\mathbf{B})$. It is not unreasonable to conjecture that

$$\lim_{n\to\infty}\frac{\sum_{G\in\mathbf{G}_n}\alpha_{f_1}(G;\mathbf{B})}{2^{\binom{n}{2}}n^2/\log n}=c$$

exists and c is probably equal to $\lim_{n\to\infty}\frac{\alpha_{f_1}(n;\mathbf{B})}{n^2/\log n}$. We can also ask the analogous question for $\beta_{f_1}(G;\mathbf{B})$.

Let $\mathbf{G}_{n,m}$ be the set of all graphs on n vertices and m edges. We can define $\alpha_f(n,m;\mathbf{H})$ to be the maximum value of $\alpha_f(G;\mathbf{H})$ where G ranges over all graphs in $\mathbf{G}_{n,m}$. In this paper we investigate $\alpha_{f_1}(n,m;\mathbf{B})$ where m is about $n^2/2e$. One could also investigate $\alpha_{f_1}(n,m;\mathbf{B})$ or $\beta_{f_1}(n,m;\mathbf{B})$. In particular, we can ask the problem of determining m so that $\alpha(n,m;\mathbf{B})$ is maximized or to find the range for m for which we have $\alpha(n,m;\mathbf{B})=o(n^2)$.

References

[1] W. G. Brown, On Graph That do not Contain a Thomson Graph, *Canad. Math. Bull.* **9** (1966), 281–285.
[2] F. R. K. Chung, On the Decomposition of Graphs, *SIAM J. on Algebraic and Discrete Methods*, **2** (1981), 1–12.
[3] F. R. K. Chung and R. L. Graham, On Multicolor Ramsey Numbers for Complete Bipartite Graphs, J. C. T. **18**, (1975), 164–169.
[4] P. Erdős and J. Spencer, Probabilistic Methods in Combinatorics, Academic Press, New York, 1974.
[5] R. L. Graham and H. O. Pollak, On the Addressing Problem for Loop Switching, *Bell Sys. Tech. Jour.*, **50** (1971), 2495–2519.
[6] F. Harary, *Graph Theory*, Addison-Wesley, New York, 1969.

BELL LABORATORIES
MURRAY HILL, NJ 07974
U.S.A.

MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY
OF SCIENCES
H 1053 BUDAPEST
REÁLTANODA U. 13–15
HUNGARY

SUNY STONY BROOK
STONY BROOK, NY
U.S.A.