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 $G_1, \ldots, G_s$, such that any edge of G is in exactly one of the $G_i$s. Let $\alpha(G; B)$ denote the minimum value of $\sum V(G_i)$ over all B-decompositions of G. Let $\alpha(n; B)$ denote the maximum value of $\alpha(G; B)$ over all graphs on $n$ vertices.

A B-covering of G is a collection of complete bipartite subgraphs $G_1, G_2, \ldots, G_s$, such that any edge of G is in at least one of the $G_i$s. Let $\beta(G; B)$ denote the minimum value of $\sum V(G_i)$ over all B-coverings of G and let $\beta(n; B)$ denote the maximum value of $\beta(G; B)$ over all graphs on $n$ vertices.

In this paper, we show that for any positive $\epsilon$, we have

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

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 $G_1, G_2, \ldots, G_s$, 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.
\[ \alpha_f(G; H) = \min_P \sum_i f(G_i) \]

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

\[ \alpha_f(n; H) = \max_G \alpha_f(G; H) \]

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 \( \alpha_f(G; P) \) is simply the thickness of \( G \). If \( F \) denotes the family of forests, then \( \alpha_f(G; F) \) 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 \( H = B \), the family of complete bipartite graphs. By a theorem in [5], the value of \( \alpha_f(n; B) \) is given by:

\[ \alpha_f(n; B) = n - 1. \]

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 \( \alpha(n) = \alpha_f(n; B) \) and \( \alpha(G) = \alpha_f(G; B) \). In particular, we show for any given \( \varepsilon \) and sufficiently large \( n \),

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

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

\[ \beta_f(G; H) = \min_P \sum_i f(G_i') \]

where \( P = \{G_1', \ldots, G_t'\} \) ranges over all \( H \)-coverings of \( G \).

It is easily seen that

\[ \beta_f(G; H) \leq \alpha_f(G; H) \]

and

\[ \beta_f(n; H) \leq \alpha_f(n; H). \]

We will show that the asymptotic growth of \( \beta_f(n; B) \) is quite similar to \( \alpha_f(n; B) \). In fact, we will obtain the same upper and lower bounds for \( \beta_f(n; B) \) as those for \( \alpha_f(n; B) \) 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.** \( \alpha(n) \geq (1 - \varepsilon) \frac{n^2}{2e \log n} \) for any given positive \( \varepsilon \) 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

\[
1 \leq a, b \leq n, \quad \frac{a+b}{ab} < \frac{1-\varepsilon}{\log n}.
\]

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

\[
\sum_{\{a,b\} \in S} \frac{n^2}{a} e^{-ab} < \sum_{\{a,b\} \in S} e^{-ab} < \sum_{\{a,b\} \in S} e^{-(\log n)^2} < n^2 e^{-(\log n)^2} < 1
\]

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 \( \mathcal{B} \)-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

\[
f(u, v) = \frac{|V(G_i)|}{|E(G_i)|}
\]

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 $G_i = K_{c,d}$, $1 \leq i \leq t$, satisfies that 
$$\frac{c + d}{cd} \geq \frac{1 - \varepsilon}{\log n}.$$ 
Thus we have 
$$f(u, v) \geq \frac{1 - \varepsilon}{\log n}$$ 
for any $\{u, v\}$ in $E(G)$. 
and 
$$\alpha(n) > \alpha(G) > \frac{(1 - \varepsilon)n^2}{2\varepsilon \log n}$$ 
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 $\rho \frac{n}{2}$ edges contains a complete bipartite graph $K_{s,t}$ as a subgraph where $t = \lfloor (1 - \varepsilon)n\rho^2 \rfloor$ and $s < \varepsilon n$ for $n$ sufficiently large.

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

$$n(\rho n - s)^2 \leq (t - 1) \cdot n^2.$$ 

However, on the other hand, we have 

$$(t - 1)n^2 < tn^2 \leq (1 - \varepsilon)n^4 + s\rho^2 < n(\rho n - s)^2$$

since $s < \varepsilon n$.

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

**Theorem 2.** For any given $\varepsilon$, we have 

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

if $n$ is large enough.

**Proof.** From Lemma 1, one can easily verify that a graph $G$ on $\rho \frac{n}{2}$ 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 $\varepsilon_1 > \frac{\log n}{\rho n}$. 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 $G_1$ be the subgraph of $G$ containing all edges of $G$ except those in $H$. Now, we find a subgraph $H_1$ isomorphic to $K_{s,t}$, and let $G_2$ be a subgraph of $G_1$ containing all edges of $G_1$ except those in $H_1$. We continue this process until we reach $G_k$ for which we cannot find a subgraph isomorphic to $K_{s,t}$. Then, we can choose any $\alpha(k)$.

This is the end of the proof for Theorem 2.
edges of $G_1$ except those in $H_1$ and continue in this fashion until only $\frac{e_2 n^2}{\log n}$ edges are left. Thus $G$ is decomposed into $H, H_1, \ldots$, together with $\frac{e_2 n^2}{\log n}$ edges and we have the following recursive relation

$$
\alpha(G) \leq s + t + \alpha(G_1).
$$

We will prove by induction that for a given $\varepsilon < e_2 < e_1, e_3 > 0$ and sufficiently large $n$ the following holds,

$$
\alpha(G) \leq (1 + e_2) \frac{n^2}{2 \log n} \int_0^\rho \log \left( \frac{1}{x} \right) dx + 2e_2 \frac{n^2}{\log n}.
$$

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

$$
\alpha(G) \leq (1 - e_2) (\log n)^3 / (\log (1/\rho))^3 + (1 + e_2) \frac{n^2}{2 \log n} \int_0^\rho \log \left( \frac{1}{x} \right) dx + 2e_2 \frac{n^2}{\log n}
$$

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

$$
(1 - e_2) (\log n)^3 / (\log (1/\rho))^3 + (1 + e_2) \frac{n^2}{2 \log n} \int_0^\rho \log \left( \frac{1}{x} \right) dx \leq
$$

$$
\leq (1 + e_2) \frac{n^2}{2 \log n} \int_0^\rho \log \left( \frac{1}{x} \right) dx
$$

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

$$
\alpha(n) \leq (1 + e_2) \frac{n^2}{2 \log n} \int_0^{1/ho} \log \left( \frac{1}{x} \right) dx + 2e_2 \frac{n^2}{\log n} \leq (1 + e) \frac{n^2}{2 \log n}
$$

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

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

**Theorem 3.**

$$
\beta_f(n; B) \geq (1 - \varepsilon) \frac{n^2}{2e \log n}
$$

for any positive $\varepsilon$ and sufficiently large $n$.  

$\gamma$
Therefore we have
\[
(1 - \varepsilon) \frac{n^2}{2e \log n} < \beta_f(n; \mathbf{B}) \leq \alpha_f(n; \mathbf{B}) < (1 + \varepsilon) \frac{n^2}{2 \log n}
\]
for any given positive \( \varepsilon \) 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, \( c_1 n^2 / \log n \) edges (or \( c_2 n^2 \) 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

\[
T(K_{t,t}; n) = O(n^{2 - \frac{1}{t}})?
\]

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

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

\[
\lim_{n \to \infty} \frac{\alpha_f(n; \mathbf{B})}{n^2 / \log n} \quad \text{or} \quad \lim_{n \to \infty} \frac{\beta_f(n; \mathbf{B})}{n^2 / \log n},
\]

obviously.

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

\[
\sum_{G \in G_n} \alpha_f(G; \mathbf{B}) = \lim_{n \to \infty} \frac{\sum_{G \in G_n} \alpha_f(G; \mathbf{B})}{2^{\binom{n}{2}} n^2 / \log n} = c
\]

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

Let \( G_{n,m} \) 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 \( G_{n,m} \). In this paper we investigate \( \alpha_f(n, m; \mathbf{B}) \) where \( m \) is about \( n^2 / 2e \). One could also investigate \( \alpha_f(n, m; \mathbf{B}) \) or \( \beta_f(n, m; \mathbf{B}) \). 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 \( \alpha(n, m; \mathbf{B}) = \omega(n^2) \).
References