MINIMAL DECOMPOSITIONS OF TWO GRAPHS INTO PAIRWISE ISOMORPHIC SUBGRAPHS

by

F. R. K. Chung
Bell Laboratories
Murray Hill, New Jersey

P. Erdös†
Mathematics Institute of the Hungarian
Academy of Sciences
Budapest, Hungary

R. L. Graham
Bell Laboratories
Murray Hill, New Jersey

S. M. Ulam
Santa Fe, New Mexico

and

F. F. Yao
Stanford University
Stanford, California

INTRODUCTION

For a graph* G with edge set $E(G)$, let $e(G)$ denote the cardinality of $E(G)$. Suppose we are given two graphs G and G' with $e(G) = e(G')$. By a U-decomposition of G and G' we mean a pair of partitions $E(G) = E_1 + \ldots + E_r$ and $E(G') = E_1' + \ldots + E_r'$ such that as graphs, E_i and E_i' are isomorphic for all i. Such decompositions always exist when G and G' have the same number of edges since we can always take each E_i and E_i' to be a single edge. The function $U(G,G')$ is defined to be the minimum value

† P. Erdös, some of this work was done while he was a consultant in Dept. 1216.

* In general, we follow the terminology of [3].

PROC. 10TH S-E CONF. COMBINATORICS, GRAPH THEORY, AND COMPUTING, pp. 3-18.

of r for which a U-decomposition of G and G' into r parts exists. A number of well-studied graph theoretic questions can be expressed in terms of U-decompositions. For example, when G' consists of $e(G)$ disjoint edges then $U(G,G')$ is just the edge-chromatic number of G and so, is always equal to δ or $\delta + 1$ where δ is the maximum degree of G (e.g., see [5], [13]). When G' consists of $e(G)$ edges incident to a single vertex then $U(G,G')$ is known as the edge-dominating number of G. Similarly, $\min_{G'} U(G,G')$ is called the thickness, arboricity or biparticity of G (see [7], [8]) where G' ranges over all planar graphs, acyclic graphs or bipartite graphs, respectively.

In this note we investigate various properties of $U(G,G')$. Of particular interest will be the quantity $U(n)$, defined by

$$U(n) = \max_{G,G'} U(G,G')$$

where G and G' each have n vertices (and, of course, $e(G) = e(G')$).

Before proceeding to the main results we first make remarks concerning notation. The set and number of vertices of a graph G will be denoted by $V(G)$ and $v(G)$, respectively. As usual, the symbols S_n, K_n and $K_{m,n}$ denote the graphs: n-star, complete graph on n vertices and complete bipartite graph on m and n vertices, respectively. The distance between x and y in G, denoted by $d_G(x,y)$, is defined to be minimum number of edges in any path joining x and y (if it exists). If no such path exists, $d_G(x,y)$ is defined to be ∞. The notation $H \subseteq G$ indicates that H is a (partial) subgraph of G; m disjoint copies of G are denoted by mG. Sometimes we write $G = G(n,e)$ to indicate that G has n vertices and e edges. Finally, $\lceil x \rceil$ denotes the "ceiling" function of x, i.e., the least integer not less than x, and $\lfloor x \rfloor$ denotes the greatest integer not exceeding x.

BOUNDS ON U(n)

The first lower bound on U(n) which is likely to occur to anyone thinking about the problem is

$$(1) \qquad U(n) \geq \left\lfloor \frac{n}{2} \right\rfloor .$$

This can be seen in a variety of ways, perhaps the simplest furnished by considering the graphs $\left\lfloor \frac{n}{2} \right\rfloor S_1$ and $S_{\left\lfloor \frac{n}{2} \right\rfloor}$ (with the appropriate number of isolated vertices shown in Fig. 1. In this case, all the E_i must be taken to be single edges.

$S_{\lfloor \frac{n}{2} \rfloor}$ $\qquad$ $\lfloor \frac{n}{2} \rfloor S_1$

Figure 1

In fact, as we shall see later, $\lfloor \frac{n}{2} \rfloor$ is the essentially best possible lower bound for U(G,G') when both G and G' are forests (i.e., acyclic). However, substantially worse examples exist when G and G' are allowed more freedom. (The reader may find it instructive to look for one before proceeding.)

A better lower bound can be obtained by considering the graphs $G = S_{3m}$ and $G' = mK_3 + S_0$. In this case

(2) $$U(G,G') = 2m,$$

the minimum U-decomposition being attained by partitioning the edges of G and G' into m S_2's and m S_1's. It follows from this that

$$(3) \qquad U(n) \geq \frac{2n}{3} + O(1).$$

As we shall see in Theorem 1, the constant $\frac{2}{3}$ is asymptotically best possible.

Concerning upper bounds for U(n), as we have noted earlier, for G = G(n,e), G' = G'(n,e), it is trivial that $U(n) \leq e \leq \binom{n}{2}$. It is almost as easy to prove the following linear upper bound on U(n):

$$(4) \qquad U(n) \leq 2n-1.$$

To see this, suppose we are given two graphs G = G(n,e) and G' = G'(n,e). Sequentially decompose them into stars by repeatedly choosing (arbitrary) nonisolated vertices $v \in G$, $v' \in G'$ and removing a S_δ centered at each of v and v' where δ is the minimum of the degrees of v and v'. At each step of this process we create at least one more isolated vertex in the remaining graphs. Thus, after at most 2n-1 steps, all the vertices in one of the graphs (and therefore, also the other graph) will be isolated and consequently, $U(G,G') \leq 2n-1$. Since G and G' were arbitrary, (4) follows.

In order to improve (4) we will have to work considerably harder. This will be done in the following theorem, which is the main result of the paper.

Theorem

(5) $$U(n) = \frac{2}{3} n + o(n).$$

Before proving (5) we first need to establish several lemmas (which are actually of independent interest).

Lemma 1. If $e(G) \geq 2ab$ then G contains either S_a or bS_1 as a subgraph.

Proof: Suppose $S_a \not\subseteq G$. Thus, $\delta(G)$, the maximum degree occurring in G, is at most a-1. Let e_1 be an arbitrary edge of G. Since $\delta(G) \leq a-1$ then e_1 is incident to at most 2a-4 other edges of G. Let e_2 be an edge which is disjoint from e_1. As before, e_2 is incident to at most 2a-4 other edges and so, e_1 and e_2 together are incident to at most 2(2a-4) other edges. Repeating this process, after $e_1, e_2, \ldots, e_k$ disjoint edges are chosen, we can continue provided $e(G) > (2a-3)k$. In particular since by hypothesis $e(G) \geq 2ab$ then G contains b disjoint edges, i.e., $bS_1 \subseteq G$ and the lemma is proved.

Lemma 2. Any two graphs $G = G(n,e)$ and $G' = G'(n,e')$ have isomorphic subgraphs with at least $\frac{ee'}{\binom{n}{2}}$ edges.

Proof: We make use of a simple application of the probability method (see [4]). Label the vertices (arbitrarily) of G and G', say, $V(G) = \{x_1,\ldots,x_n\}$, $V(G') = \{x'_1,\ldots,x'_n\}$. Let Λ denote the set of all 1-1 mappings of $V(G)$ onto $V(G')$. Thus, $|\Lambda| = n!$. For given edges $y \varepsilon E(G)$, $y' \varepsilon E(G')$, there are exactly $2(n-2)!$ elements $\lambda \varepsilon \Lambda$ which map y onto y'. Hence, if we define

$$i_\lambda(y,y') = \begin{cases} 1 \text{ if } \lambda \text{ maps } y \text{ onto } y', \\ 0 \text{ otherwise,} \end{cases}$$

then

$$\sum_{\lambda \varepsilon \Lambda} \sum_{y,y'} i_\lambda(y,y') = \sum_{y,y'} \sum_{\lambda \varepsilon \Lambda} i_\lambda(y,y')$$

$$= \sum_{y,y'} 2(n-2)! = 2ee'(n-2)!$$

Since $|\Lambda| = n!$ then for some $\lambda_0 \varepsilon \Lambda$

$$\sum_{y,y'} i_{\lambda_0}(y,y') \geq \frac{2ee'(n-2)!}{n!} = \frac{ee'}{\binom{n}{2}}.$$

The edges of G mapped onto edges of G' by λ_0 form the subgraph of the required size. ∎

We now begin the proof of (5). Starting with the given graphs $G = G(n,e)$ and $G' = G'(n,e)$, where n is a large integer, repeatedly apply Lemma 2, removing the large common subgraphs which are guaranteed by the lemma. Thus, if after the kth step we have $G_k = G_k(n,e_k)$ and $G_k' = G_k'(n,e_k)$ remaining, then the lemma implies that we can remove isomorphic subgraphs having at least $e_k^2/\binom{n}{2}$ edges. This leaves graphs G_{k+1} and G_{k+1}' each having

$$(6) \qquad e_{k+1} \leq e_k - e_k^2/\binom{n}{2}$$

edges, (where we let $e_0 = e$). Let α_k denote $e_k/\binom{n}{2}$; hence, $0 \leq \alpha_k \leq 1$ and (6) can be rewritten as

$$(6') \qquad \alpha_{k+1} \leq \alpha_k - \alpha_k^2 \equiv f(\alpha_k),\ k \geq 0.$$

Note that $y = f(x)$ is just a parabola with a maximum of 1/4 at $x = 1/2$.

Suppose for some k that $\alpha_k < 1/k \leq 1/2$. Since $f(x)$ is monotone for $0 \leq x \leq 1/2$, we have

$$\alpha_{k+1} \leq f(\alpha_k) \leq f(1/k) = \frac{1}{k} - \frac{1}{k^2} = \frac{k-1}{k^2} < \frac{1}{k+1}.$$

But $\alpha_1 = f(\alpha_0) \leq 1/4$ so by induction

(7) $$\alpha_k < 1/k, \ k \geq 1.$$

Translating (7) back to e_k we obtain

(7') $$e_k < \binom{n}{2}/k, \ k \geq 1.$$

We note that this result can already be used to improve the trivial bound of 2n-1 on U(n). For, taking $k = \lfloor n/\sqrt{2} \rfloor$ in (7') we have $e_k < n/\sqrt{2}$ which implies that the decompositions of the remaining graphs G_k and G_k' can then be completed (one edge at a time) in at most $n/\sqrt{2}$ steps. Thus, the U-decomposition requires altogether $n/\sqrt{2}$ steps, and we have shown that $U(n) \leq n\sqrt{2}$.

Coming back to the proof of (5), we apply (7') with $k = \frac{n}{C}$ for a large number C to be specified later. Let $H = G_k$ and $H' = G_k'$ denote the two graphs remaining at this stage. Thus, $e(H) = e(H') \leq \frac{1}{2}Cn$.

We next sequentially remove all common S_m's and mS_1's from H and H' with $m = \lceil \log n \rceil$. We continue until the remaining graphs, which we denote by J and J', respectively, contain no common S_m's and mS_1's. Without loss of generality we may assume

$$S_m \not\subseteq J, \ mS_1 \not\subseteq J'.$$

Note that in going from H to J we used at most $e(H)/m \leq \frac{Cn}{2 \log n}$ steps since each step removes m edges. If $e(J) \leq n/\log n$ then we are done since we can complete the U-decomposition in at most $\frac{n}{\log n}$ steps, the whole process requiring at most

$$\frac{n}{C} + \frac{Cn}{2 \log n} + \frac{n}{\log n} = o(n)$$

steps for large C and large n. Hence, we may assume $e(J) > \frac{n}{\log n}$.

It now follows from Lemma 1 that

$$(8) \qquad \delta(J') \geq e(J')/2m > n^{2/3}$$

for large n since $mS_1 \not\subseteq J'$.

Let $V' = \{v_1, v_2, \ldots, v_r\}$ denote the set of vertices v of J' which have $\deg(v) \geq \delta(J') - 1$. In other words, V' consists of all vertices having maximum degree or maximum degree minus one in J'. By (8) it follows that $r \leq Cn^{1/3}$. In J, choose a set $W = \{w_1, w_2, \ldots, w_r\}$ of r vertices with the properties:

(i) for all $i \neq j$, $d_J(w_i, w_j) \geq 3$;

(ii) if $x \notin W$ and $d_J(x, w_j) \geq 3$ for all j then $\deg(x) \leq \min_i\{\deg(w_i)\}$.

To see that such a set W exists, observe that since $\delta(J) < m$, for each vertex v of J there are at most

m^2 vertices x with $d_J(v,x) \leq 2$. Thus, we can certainly find a set W_0 with at least $v(J)/m^2 > n/2 \log^2 n > r$ vertices. If W_0 doesn't satisfy (ii), successively replace vertices in it by outside vertices of larger degree until it does.

There are now two possibilities:

(a) $\min\{\deg(w_i)\} \geq 2$. In this case the common subgraph X we remove from J and J' is

$$X = S_{\deg(w_1)} + S_{\deg(w_2)} + \dots + S_{\deg(w_r)},$$

forming $\bar{J}$ and $\bar{J}'$, respectively. This is certainly possible in J since we simply form the r stars of X at the corresponding r vertices in W. On the other hand it is not difficult to see that X also occurs in J', with the $S_{\deg(w_i)}$ centered at the v_i, $1 \leq i \leq r$, since $\deg(v_i) \geq \delta(J') - 1 > n^{2/3}$, $r < Cn^{1/3}$ and $\deg(w_i) < m = \lceil \log n \rceil$ for large n. Let I(Y) denote the number of isolated vertices of a graph Y. Then

$$I(\bar{J}) \geq I(J) + 1 \text{ and } \delta(\bar{J}') \leq \delta(J') - 2$$

and so,

$$n - I(\bar{J}) + \delta(\bar{J}') \geq n - I(J) + \delta(J') + 3.$$

(b) $\min_i\{\deg(w_i)\} = 1$. Since $\delta(J) < m$ there are at most $m^2 < \sqrt{n}$ edges within distance 2 from some vertex of W. Hence, at least $\frac{n}{\log n} - r\sqrt{n} > \frac{n}{2 \log n}$ edges

$\{x,y\}$ of J have both endpoints x and y a distance of at least 3 from every $w_i \varepsilon W$. By the definition of W, this implies that the edge $\{x,y\}$ is <u>isolated</u>. In this case the common subgraph X we remove from J and $\overline{J}$ is $X = rS_1$. In J, we form X from r isolated edges, thereby isolating at least <u>two</u> new vertices in forming $\overline{J}$. In J', we form $\overline{J}' = J' - X$ by removing disjoint edges incident to each of the $v_i \varepsilon V$. Therefore,

$$\begin{aligned} n-I(\overline{J}) + \delta(\overline{J}') &\geq n-I(J) + 2 + \delta(J') + 1 \\ &= n-I(J) + \delta(J') + 3. \end{aligned}$$

Thus, in either case $n-I(J) + \delta(J')$ increases by at least 3. Since at the beginning of this reduction, $I(J) \geq 0$, $\delta(J') \leq n$ then $n-I(J) + \delta(J') \leq 2n$. Thus, we can apply this reduction until the remaining graphs $\hat{J}$ and $\hat{J}'$ have $e(\hat{J}) = e(\hat{J}') \leq \frac{n}{\log n}$. This part of the reduction requires at most $\frac{2n}{3}$ steps.

The final decomposition of $\hat{J}$ and $\hat{J}'$ uses at most $e(J) \leq \frac{n}{\log n}$ steps. Thus, the complete U-decomposition of G and G' requires at most

$$\frac{n}{C} + \frac{Cn}{2 \log n} + \frac{2n}{3} + \frac{n}{\log n} < \left(\frac{2}{3} + \varepsilon\right)n$$

steps for any fixed $\varepsilon > 0$ provided C and n are sufficiently large. Thus,

(9) $$U(n) \leq \frac{2}{3} n + o(n).$$

Equation (5) now follows from (3) and (9) and the Theorem is proved. ∎

VARIATIONS AND EXTENSIONS

If G and G' are both restricted to be bipartite graphs then a somewhat stronger bound on U(G,G') holds. In particular, if $U_B(n)$ denotes

$$\max\{U(G,G'):G,G' \text{ bipartite}, v(G) = v(G') = n, e(G) = e(G')\}$$

then it can be shown using arguments quite similar to those used for (5) that

(10) $$U_B(n) = \frac{n}{2} + o(n).$$

(The example given in Fig. 1 shows that $\frac{n}{2}$ is a lower bound for $U_B(n)$).

A natural extension of our problem which suggests itself is the generalization to more than two graphs. Specifically, for a given integer k, define

$$U_k(n) = \max\{U(G_1,\ldots,G_k):v(G_1) = \ldots = v(G_k) = n, e(G_1) = \ldots = e(G_k)\}$$

where $\bar{U}(G_1,\ldots,G_k)$ denotes the least integer r such that there exist partitions $E(G_i) = E_{i,1} + \ldots + E_{i,r}$ having for each j all $E_{i,j}$ isomorphic, $1 \leq i \leq r$. Quite surprisingly, it turns out that

$$U_k(n) = \frac{3}{4} n + o(n).$$

for any fixed $k \geq 3$. The proof, however, is significantly more complicated than our proof of (5) and will be dealt with in a later paper.

From an algorithmic point of view, it is of interest to know how hard it is to determine $U(G,G')$ in general. The question "Is $U(G,G') = 1$?", also known as the <u>graph isomorphism</u> problem, has been extensively investigated recently [12]. It is not known to be NP-complete and in fact, opinions on whether or not it is are mixed (see [1] or [6] for relevant definitions). As mentioned previously, the value $U(G,e(G)S_1)$ is just the edge-chromatic number $\chi_e(G)$ of G. However, the determination of $\chi_e(G)$ is also not known to be NP-complete. On the other hand, the determination of $U(G,S_{e(G)})$ is known to be NP-complete (see [10]). In fact, one of the authors (F.Y. [14]) has very recently shown that the question "Is $U(G,G') = 2$?" is NP-complete. It would be interesting to know if that still held when G and G' were restricted to be trees (for example).

Of course, most of the preceding questions can be raised for directed graphs, hypergraphs or more generally, for other combinatorial, algebraic or geometrical structures. A particular example of this which has received a fair amount of attention during the last 75 years is the concept of equidecomposability of planar polygonal figures. It was shown by Bolyai and Gerwien and also by Hilbert (see [11]) that any two polygonal regions P, P' in the plane having the same area could be decomposed into finitely many (connected) regions, say $P = P_1 + \dots + P_r$, $P' = P_1' + \dots + P_r'$, such that P_i and P_i' are congruent for all i (where we ignore what happens to boundary points of the P_i). The determination of the minimum number H(P,P') of regions for which this is possible is an extremely difficult and largely untouched area. For example, if P(k) denotes the regular k-gon with area 1, it is known that $H(P(5),P(4)) \leq 6$. There seem to be no methods currently available for proving that H(P(5),P(4)) = 6, if indeed it does. In fact, there are no reasonable bounds on H(P(m),P(n)) generally. As is well known, the analogous theorem does not hold for equal volume polyhedra in 3-space (to show this was exactly Hilbert's 3rd problem [9]). For an interesting account of this subject the reader is referred to the recent book [2].

REFERENCES

1. A. V. Aho, J. E. Hopcroft and J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Mass. (1974).

2. V. G. Boltianskii, Hilberts Third Problem, V. H. Winston and Sons, Washington, D.C. (1978).

3. J. A. Bondy and U. S. R. Murty, Graph Theory With Applications, American Elsevier, New York (1976).

4. P. Erdös and J. Spencer, Probabilistic Methods in Combinatorics, Academic Press, New York (1974).

5. S. Fiorini, A bibliographic survey of edge-coloring, J. of Graph Theory, 2 (1978), 93-106.

6. M. R. Garey and D. S. Johnson, Computers and Intractability, W. H. Freeman and Co., San Francisco (1979).

7. F. Harary, Graph Theory, Addison-Wesley, Reading, Mass. (1969).

8. F. Harary, D. Hsu and Z. Miller, The biparticity of a graph, J. of Graph Theory, 1 (1977), 131-134.

9. D. Hillbert, Mathematical Problems, lecture delivered before the International Congress of Math. in Paris in 1900, translated by M. W. Newson, Bull. Amer. Math. Soc. 8 (1902), 437-479.

10. R. M. Karp, Reducibility Among Combinatorial Problems, in Complexity of Computer Computations, R. E. Miller and J. W. Thatcher (eds.), Plenum Press, New York, (1972).

11. H. Lindgren, Recreational Problems in Geometric Dissections and How to Solve Them, Dover, New York, (1972).

12. R. C. Read and D. G. Corneil, The graph isomorphism disease, J. of Graph Theory, 1 (1977), 339-364.

13. V. G. Vizing, On an estimate of the chromatic class of a p-graph, Diskret. Analiz., 3 (1964), 25-30.

14. F. Yao, (to appear).