An Extremal Problem for Paths in Bipartite Graphs

A. Gyárfás
HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, HUNGARY

C. C. Rousseau
R. H. Schelp
MEMPHIS STATE UNIVERSITY
MEMPHIS, TENNESSEE 38152

ABSTRACT

A formula is found for the maximum number of edges in a graph $G \subseteq K(a, b)$ which contains no path P_{2l} for $l > c$. A similar formula is found for the maximum number of edges in $G \subseteq K(a, b)$ containing no P_{2l+1} for $l > c$. In addition, all extremal graphs are determined.

1. INTRODUCTION

Various extremal problems for paths and cycles in graphs were considered by Erdös and Gallai in [3]. Among their several conclusions is the following result.

Theorem (Erdös and Gallai). Let G be a graph of order n which contains no path with more than k vertices. Then

$$q(G) \leq \frac{1}{2}(k - 1)n$$

and equality holds iff each component of G is a complete graph of order k.

It is natural to seek similar results by considering the extremal problem for paths in which the graph G is further restricted in some way. For example, in [3] Erdös and Gallai also consider the case in which G is a *connected* graph of order n containing no path with more than k vertices. In this paper we consider the extremal problem for paths in *bipartite* graphs.

Journal of Graph Theory, Vol. 8 (1984) 83–95

CCC 0364-9024/84/010083-13$04.00

Let us agree to call a path *even* or *odd* according to the parity of the number of its vertices. In addressing the extremal problem for paths in bipartite graphs, it is appropriate to take into account the fact that paths of different parity are qualitatively different, since the parity determines whether the endvertices of the path are in the same or opposite parts of the graph. Accordingly, we shall formulate two separate problems, one for each parity. The terminology in which these problems are formulated now follows.

The bipartite graph with parts X and Y and with edge set E is denoted $G(X, Y, E)$. Let a, b, and c be specified natural numbers with $a \leqslant b$. A bipartite graph $G(X, Y, E)$ with $|X| = a$ and $|Y| = b$ which contains no even path P_{2l} for $l > c$ will be called *feasible* and the collection of all feasible graphs will be denoted $F_0(a, b, c)$. Among the feasible graphs, those with the greatest possible number of edges will be called *extremal* and the collection of all extremal graphs will be denoted $E_0(a, b, c)$. The number of edges $q(G)$ of an extremal graph G is given by $f_0(a, b, c)$. The extremal problem for even paths, then, is to calculate f_0 and to determine the collection of extremal graphs E_0. In the extremal problem for odd paths, a graph will be called feasible if it contains no odd path P_{2l+1} for $l > c$. Now our previous notation is essentially recycled, with the feasible graphs, extremal graphs, and the extremal number of edges being $F_1(a, b, c)$, $E_1(a, b, c)$, and $f_1(a, b, c)$, respectively. Again, the problem is to calculate f_1 and to determine the collection of extremal graphs E_1.

In addition to the special notation just introduced, there will be exceptional symbolism used at certain points in the paper. Such notation will be explained at the appropriate juncture. For the most part, however, our terminology and notation will be standard and follow either [1] or [2].

2. DESCRIPTION OF THE EXTERNAL GRAPHS

Bipartite graphs considered in this paper will be of the form $G(X, Y, E)$ where $X = \{x_1, x_2, \ldots, x_a\}$, $Y = \{y_1, y_2, \ldots, y_b\}$, and $a \leqslant b$. The edge set of such a graph may be specified by means of the $a \times b$ *matching matrix* $M(G) = [m_{ij}]$ defined by setting

$$m_{ij} = \begin{cases} 1 & \text{if } x_i y_j \in E(G), \\ 0 & \text{if } x_i y_j \notin E(G). \end{cases}$$

This description is particularly useful as a means of specifying the members of the extremal classes $E_0(a, b, c)$ and $E_1(a, b, c)$. In particular, let us focus on those graphs G for which $M(G)$ can be written in partitioned

TABLE I.

Graph	Range	s	t^a	M_{11}	M_{12}	M_{21}	M_{22}
G_{01}	$a \leqslant c$	a	b	1	···	···	···
G_{02}	$c < a \leqslant 2c$	c	b	1	···	0	···
G_{03}	$a = 2c$	c	*	1	0	0	1
G_{04}	$a > 2c$	c	$b - c$	1	0	0	1

[a]* = arbitrary.

form (using at most four blocks) as

$$M(G) = \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right],$$

where each block in the partitioned matrix is either a matrix of all 1's or else a matrix of all 0's. Such a matrix is completely specified by giving the size of M_{11} ($s \times t$) and identifying each M_{ij} as a block of all 1's or all 0's. The obvious interpretation is made in case $s = a$ or $t = b$. Table I defines those graphs which will be shown to be the extremal ones for even paths.

Similarly, Table II defines those graphs which will be shown to be the extremal ones for odd paths when $c > 2$.

TABLE II.

Graph	Range	s	t	M_{11}	M_{12}	M_{21}	M_{22}
G_{11}	$a \leqslant c$ or $a = b = c + 1$	a	b	1	···	···	···
G_{12}	$a = c + 1$, $b = c + 2$	$c + 1$	$c + 1$	1	0	···	···
G_{13}	$a = b = c + 2$	$c + 1$	$c + 1$	1	0	0	1
G_{14}	$b > a = c + 1$ or $c + 1 < a < 2(c + 1)$	c	$b - 1$	1	1	0	1
G_{15}	$a = 2c + 1$ or $a = b = 2(c + 1)$	$c + 1$	$c + 1$	1	0	0	1
G_{16}	$b > a = 2(c + 1)$ or $a > 2(c + 1)$	c	$b - c$	1	0	0	1

TABLE III.

Range	$f_0(a, b, c)$
$a \leqslant c$	ab
$c < a < 2c$	bc
$a \geqslant 2c$	$(a + b - 2c)c$

The fact that Table I defines the extremal class $E_0(a, b, c)$ means that f_0 is given as listed in Table III. Similarly, the fact that for $c > 2$ the extremal class $E_1(a, b, c)$ is as given in Table II has the consequences concerning f_1, shown in Table IV.

3. PROOF OF THE EXTREMAL RESULTS FOR EVEN PATHS

As indicated in the last section, the extremal results for even paths are summarized as follows.

Theorem 1. Let $G(X, Y, E)$ be a bipartite graph with $|X| = a$ and $|Y| = b$ $(a \leqslant b)$ which contains no even path P_{2l} for $l > c$. Then $q(G) \leqslant f_0(a, b, c)$ where f_0 is defined in Table III. Furthermore, equality holds iff G is one of the graphs listed in Table I.

The proof of this theorem will be postponed until appropriate preparation has been made. To begin with, it is helpful to know the following fact.

Lemma 1. The statement of Theorem 1 is true in case each connected component of G is a complete bipartite graph.

Proof. Suppose that the (nontrivial) connected components of G are complete bipartite graphs with parts $X_j \subseteq X$, $Y_j \subseteq Y$, $j = 1, \ldots, n$. Isolated

TABLE IV.

Range	$f_1(a, b, c)$
$a \leqslant c$	ab
$a = b = c + 1$	$(c + 1)^2$
$b > a = c + 1$ or $c + 1 < a < 2(c + 1)$	$a + (b - 1)c$
$a = b = 2(c + 1)$	$2(c + 1)^2$
$b > a = 2(c + 1)$ or $a > 2(c + 1)$	$(a + b - 2c)c$

vertices, if any, will taken to be elements of X; an easy argument shows that no graph with isolates in Y is extremal. Consequently, if $|X_j| = a_j$ and $|Y_j| = b_j$, $j = 1, \ldots, n$, we must have

$$\sum_{j=1}^{n} a_j \leqslant a, \tag{1}$$

$$\sum_{j=1}^{n} b_j = b, \tag{2}$$

$$\min(a_j, b_j) \leqslant c, \qquad j = 1, \ldots, n. \tag{3}$$

Our problem is purely arithmetical. We seek the maximum possible value of

$$q(G) = \sum_{j=1}^{n} a_j b_j$$

subject to constraints (1)–(3). If $n = 1$ the maximum value of q is clearly $\min(a, c)b$ and it is uniquely realized by graph G_{01} or G_{02} of Table I (depending on the values of a and c). Henceforth, assume that $n > 1$ and that G realizes the maximum possible value of q. The following property is implied by the assumed maximality of q.

$$\text{If } a_i, b_k > c \text{ then } a_k = b_i = c \text{ and } a_j, b_j \geqslant c \text{ for all } j. \tag{$*$}$$

To see this, first note that (3) implies that $a_k, b_i \leqslant c$. Now suppose that $b_i < c$ and consider the effect of increasing b_i by 1 and simultaneously decreasing b_k by 1. Then (1)–(3) would still be satisfied and q would be increased by $a_i - a_k$, contrary to its assumed maximality. It follows that $b_i = c$. Now with $b_i = c$, suppose that $b_j < c$ for some j and consider the effect of increasing a_i by 1 and simultaneously decreasing a_j by 1. Then (1)–(3) would still be satisfied and q would be increased by $b_i - b_j$, contrary to its assumed maximality. Consequently, $b_j \geqslant c$ for all j. The same arguments show that $a_k = c$ and $a_j \geqslant c$ for all j.

Since $n > 1$, it follows that $q(G) < ab$ so we may assume that $a > c$. In the remainder of the argument, ($*$) is applied in two cases.

Case 1. $c < a \leqslant 2c$. If $a_j \leqslant c$ for all j, then the conclusion $q(G) \leqslant bc$ is immediate. Furthermore, it is easy to see that equality occurs iff $n = 2$, $a = 2c$, and $a_1 = a_2 = c$. Thus we find the graph G_{03}. The same argument applies under the assumption that $b_j \leqslant c$ for all j. The alternative is to assume that there are indices i and k such that $a_i, b_k > c$. But then ($*$) implies that $a_j \geqslant c$ for all j. This together with $a_i > c$ contradicts the fact that $a_1 + \cdots + a_n \leqslant a \leqslant 2c$.

Case 2. $a > 2c$. As in the previous argument, we need only consider the case where there are indices i and k with $a_i, b_k > c$. Otherwise, $q(G) \leq bc < (a + b - 2c)c$. Now by combining condition (3) with the consequence of (*), we realize that for every j ($j = 1, \ldots, n$) either $a_j = c$ or $b_j = c$. It follows that $q(G) \leq (a + b - nc)c \leq (a + b - 2c)c$. Furthermore, equality occurs iff $n = 2$ and $a_1 = b_2 = c$, $a_2 = a - c$, $b_1 = b - c$. In other words, equality occurs iff $G \simeq G_{04}$. ▮

The following special case will be needed as well.

Lemma 2. The statement in Theorem 1 is true in case $a = b = c + 1$.

Proof. With $|X| = |Y| = c + 1$ suppose that $G(X, Y, E)$ has at least $c(c + 1)$ edges and contains no P_{2l} for $l > c$. We wish to prove that $q(G) = c(c + 1)$ and that G is one of the graphs listed in Table I. The proof is by induction on c. For $c = 1$ it is easy to see that $q(G) = 2$ and that the two edges of G may either be adjacent or not. Thus there are two extremal graphs; this is in accordance with Table I. (In this special case $a = c + 1 = 2c$.) Now suppose that $c > 1$ and note that since G is not complete bipartite there are two adjacent vertices with degree sum at most $2c + 1$. Without loss of generality, assume $x_a y_b \in E(G)$ and $d(x_a) + d(y_b) \leq 2c + 1$. Then $H = G - \{x_a, y_b\}$ has at least $c(c - 1)$ edges. It is obvious that H is not Hamiltonian, for then the absence of a P_{2l} for $l > c$ excludes all $2c$ edges between $\{x_a, y_b\}$ and $V(H)$. (Recall that G is missing at most $c + 1$ edges.) Also, H cannot contain a Hamiltonian path. To see this, suppose that $(x_1, y_1, \ldots, x_c, y_c)$ is a path in H and note that the following $c + 2$ edges are then excluded from G: $x_1 y_b$, $x_a y_c$, $x_1 y_c$, and either $x_a y_j$ or $x_{j+1} y_b$ ($j = 1, \ldots, c - 1$). Therefore, by the induction hypothesis, H is a graph from Table I. In particular, one finds that $H \simeq K(c - 1, c) \cup K(1)$, or, in the special case of $c = 2$, $H \simeq 2K(1, 1)$. (The first of these is the graph G_{02} with parameters $a' = b' = c$, $c' = c - 1$). Now it is easy to see that the only way to restore x_a, y_b and their incident edges and not produce P_{2l} with $l > c$ is for the isolated vertex of H to be isolated in G as well. Consequently, $G \simeq G_{02}$ with $a = b = c + 1$. ▮

The following lemma requires the most technical proof to be found in this paper. At the same time, it may be the most essential result of the paper, the key result without which the (inductive) proof of Theorem 1 falls apart.

Lemma 3. With $c + 1 \leq a \leq 2c + 1$ let $G(X, Y, E)$ be a connected bipartite graph with $|X| = |Y| = a$ in which $d(x) + d(y) \geq c + 1$ for every $x \in X$ and $y \in Y$. Then G contains a path P_{2l} with $l > c$.

Proof. We shall assume that G contains no path P_{2l} with $l > c$ and show that this assumption leads to a contradiction. Consider all possible partitions $X = (X_1, X_2, \ldots, X_m, X_R)$, $Y = (Y_1, Y_2, \ldots, Y_m, Y_R)$ where (i) $|X_j| = |Y_j|$, $j = 1, 2, \ldots, m$; (ii) the bipartite graph $G_j(X_j, Y_j, E_j)$ induced in G by (X_j, Y_j) has a Hamiltonian path $(\alpha_j, \ldots, \omega_j)$ where, by convention, $\alpha_j \in X_j$ and $\omega_j \in Y_j$, $j = 1, 2, \ldots, m$; (iii) the bipartite graph induced in G by (X_R, Y_R) is empty. Let $|X_j| = |Y_j| = n_j$ and note that, by assumption, $n_j \leqslant c$, $j = 1, 2, \ldots, m$. Let $N = n_1 + n_2 + \cdots + n_m$. Finally, let $R = X_R \cup Y_R$. ▮

Let us introduce an ordering of the sequences $(n_1, n_2, \ldots, n_m)$ by first ordering with respect to N and then lexicographically. Thus $(n_1, n_2, \ldots, n_m) > (k_1, k_2, \ldots, k_l)$ if (i) $n_1 + \cdots + n_m > k_1 + \cdots + k_l$ or (ii) $n_1 + \cdots + n_m = k_1 + \cdots + k_l$ and $n_j > k_j$ at the first place where the two sequences disagree. Select a partition corresponding to the maximum element in this ordering. The assumed maximality has the following obvious consequences.

(A) It is forbidden for both α_i and ω_i to be adjacent to vertices in R.

(B) Neither α_i nor ω_i is adjacent to any vertex of $V(G_j)$ for $j > i$.

(C) If $v \in R$ is adjacent to either α_i or ω_i then v is not adjacent to any vertex of $V(G_j)$ for $j > i$.

(D) None of the first n_j vertices of X_i (in the sense of the path which begins at α_i and ends at ω_i) can be adjacent to ω_j for $j > i$.

(E) If $j > i$ and *some* vertex of X_i is adjacent to ω_j then none of the first n_j vertices of X_i are adjacent to ω_i.

Consider G_1 under the assumption that neither α_1 nor ω_1 is adjacent to any vertex in R. Note that since $n_1 \leqslant c$ and $a \geqslant c + 1$, $G_1 \neq G$. Then, in view of (B) and the fact that $d(\alpha_1) + d(\omega_1) \geqslant c + 1$, an easy argument shows that G_1 is Hamiltonian. Now any vertex in X_1 can take the place of α_1 and any vertex in Y_1 can take the place of ω_1. Consequently, the assumption that neither α_1 nor ω_1 is adjacent to a vertex in R leads to the conclusion that G is disconnected and so must be rejected. In particular, we conclude that $R \neq \emptyset$. In view of (A), we may now assume that α_1 is adjacent to some vertex in Y_R but that ω_1 is not adjacent to any vertex in X_R.

Let us define the index set I by the condition that $i \in I$ if there exists a subsequence of $G_i, G_{i+1}, \ldots, G_m$ which yields an *alternating path* from α_i to some vertex in X_R in the following sense. The path must alternate segments of the original Hamiltonian paths with edges of the form uv where $u \in Y_j$, $v \in X_k$, and G_k is the immediate successor of G_j in the subsequence. The final edge joins a vertex in one of the G's to a vertex in X_R. Finally, in traversing the path starting with α_i the segments of the Hamiltonian paths are traversed in the sense originally described, i.e., $(\alpha_j, \ldots, \omega_j)$. Since G is connected, it follows that $I \neq \emptyset$. With this in place, we are ready to add an observation to our list of consequences of maximality.

(F) If $i \in I$ then α_i is not adjacent to any vertex in Y_R.

In view of our earlier conclusion concerning α_1, it follows from (F) that $1 \notin I$. Let l denote the *least member* of I. We are now in position to make a sequence of observations related to the special nature of l. The first of these follows immediately from facts (B) and (F).

(1) α_l is not adjacent to any vertex outside of Y_l.

A similar conclusion may be drawn concerning ω_l, but proof is required.

(2) ω_l is not adjacent to any vertex in X_i for $i \neq l$.

To prove this, we begin by recalling (B) and so noting that (2) is true for $i > l$. Suppose that ω_l is adjacent to $x \in X_i$ for some $i < l$. Let $j \leqslant l$ be the *least* index such that ω_j is adjacent to a vertex in X_r for some $r < j$ and now consider the vertex ω_r. By the minimality of j it follows that ω_r is not adjacent to any vertex in X_k for $k < r$. Since $r < j \leqslant l$, our choice of l ensures that ω_r is not adjacent to any vertex in X_R. Finally, these facts together with properties (B) and (E) show that $d(\omega_r) \leqslant n_r - n_j$. It follows from fact (1) that $d(\alpha_l) \leqslant n_l$ so that $d(\alpha_l) + d(\omega_r) \leqslant n_l + n_r - n_j \leqslant n_r \leqslant c$, contrary to the hypothesis of the lemma. Thus, we have established (2).

(3) G_l is not Hamiltonian.

Suppose that G_l is Hamiltonian. Then any vertex in X_l can play the role of α_l and any vertex in Y_l can play the role of ω_l. Consequently, the only edges connecting G_l to the rest of G are between Y_l and X_R. It is now apparent that a path connecting a vertex of G_l with one in G_1 is impossible without violating either (C) or our choice of l. Therefore, we must conclude that G_l is not Hamiltonian. (It should be noted that the same argument rules out the possibility that $n_l = 1$.)

Let $x = \alpha_l$ and $y = \omega_l$. Since G_l is not Hamiltonian but $d(x) + d(y) \geqslant c + 1 \geqslant n_l + 1$, facts (1) and (2) imply that y must be adjacent to some vertex x' in X_R. Now x' can play the same role as did α_l and (since $n_l \neq 1$) there is a vertex $y' \neq y$ in Y_l to play the role of ω_l. Note that fact (3) means that no vertex in X_R is adjacent to both y and y'. Thus, if y is adjacent to s vertices in X_R and $|X_R| = r$ then y' is adjacent to at most r-s vertices of X_R. In view of fact (3), it now follows that $d(x) + d(y) + d(x') + d(y') \leqslant 2n_l + r \leqslant a$. Since $a \leqslant 2c + 1$ this is a clear violation of the hypothesis of the lemma. Thus, we have reached the desired contradiction and the lemma is proved. ∎

We are now prepared to prove the main result of this section.

Proof of Theorem 1. Let $f_0(a, b, c)$ be as defined in Table III. Suppose that $G \in F_0(a, b, c)$ and that $q(G) \geq f_0(a, b, c)$. We wish to prove that G is one of the graphs listed in Table I. The proof will be by induction on $a + b$.

Suppose that G is disconnected and let the connected components of G be $G_j \subseteq K(a_j, b_j)$, $j = 1, 2, \ldots, m$. Since $a_j + b_j < a + b$, $j = 1, 2, \ldots, m$, it follows from the induction hypothesis that $q(G_j) \leq f_0(a_j, b_j, c)$ with equality iff G_j is one of the graphs of $E_0(a_j, b_j, c)$ listed in Table I. Let H be the graph obtained from G by replacing each G_j by $H_j \in E_0(a_j, b_j, c)$. Then each connected component of H is complete bipartite. Invoking Lemma 1, we conclude that $q(G) = q(H) = f_0(a, b, c)$. This equality forces $q(G_j) = f_0(a_j, b_j, c)$, $j = 1, 2, \ldots, m$, and makes the only acceptable conclusion to be that each G_j is complete bipartite. Now the desired conclusion follows from Lemma 1.

Now suppose that G is connected. Note that the result is trivial in case $a \leq c$ and that for $a > c$ all graphs in Table I are disconnected. Hence it is our aim to show that the assumption of a connected feasible graph G with $q(G) \geq f_0(a, b, c)$ leads to a contradiction. There are three cases to consider, but in each case the nature of the argument is the same. By deleting prescribed vertices from G we obtain a graph $H \in F_0(a', b', c)$, $a' + b' < a + b$, with $q(H) \geq f_0(a', b', c)$. By the induction hypothesis, H must be a graph listed in Table I. Now the reader is asked to observe that restoration of the deleted vertices to produce a *connected,* feasible graph G with $q(G) \geq f_0(a, b, c)$ is impossible. Since the case of $a = b = c + 1$ is settled in Lemma 2, the cases for us to consider are as follows.

$\boldsymbol{a > 2c}$. Note that there exists a vertex $v \in V(G)$ with $d(v) \leq c$, for, otherwise, G contains a path of order $2(c + 1)$. Let $H = G - \{v\}$ and note that $q(G) \geq f_0(a, b, c) = (a + b - 2c)c$ implies that $q(H) \geq (a + b - 1 - 2c)c = f_0(a', b', c)$, where either $a' = a$, $b' = b - 1$ or $a' = a - 1$, $b' = b$.

$\boldsymbol{c + 1 \leq a \leq 2c}$ **and** $\boldsymbol{a < b}$. In this case, there exists a vertex $y \in Y$ with $\boldsymbol{d(y) \leq c}$. Otherwise, it follows from a theorem of Jackson [4] that G contains every even cycle C_{2k} for $2 \leq k \leq c + 1$ and so a path of order $2(c + 1)$. Let $H = G - \{y\}$ and note that $q(G) \geq f_0(a, b, c) = bc$ implies that $q(H) \geq (b - 1)c = f_0(a', b', c)$ where $a' = a$, $b' = b - 1$.

$\boldsymbol{c + 1 < a \leq 2c}$ **and** $\boldsymbol{a = b}$. Lemma 3 implies that there exist vertices $x \in X$ and $y \in Y$ with $d(x) + d(y) \leq c$. Let $H = G - \{x, y\}$ and note that $H \in F_0(a', a', c)$ (where $a' = a - 1$ satisfies $c \leq a' < 2c$) and that $q(H) \geq a'c = f_0(a', a', c)$.

Now with the conclusions that for $a > c$ there exists no connected extremal graph, the proof of Theorem 1 is complete. ∎

4. PROOF OF THE EXTREMAL RESULTS FOR ODD PATHS

Our purpose is to prove the following analog of Theorem 1.

Theorem 2. Let $G(X, Y, E)$ be a bipartite graph with $|X| = a$ and $|Y| = b$ $(a \leqslant b)$ which contains no odd path P_{2l+1} for $l > c > 2$. Then $q(G) \leqslant f_1(a, b, c)$ where f_1 is defined in Table IV. Furthermore, equality holds iff G is one of the graphs listed in Table II.

Throughout the discussion of this section, $G \in F_1(a, b, c)$ will be taken to have at least $f_1(a, b, c)$ edges. In view of the fact that $f_1(a, b, c) \geqslant f_0(a, b, c)$, Theorem 1 shows that G must be one of the extremal graphs listed in Table 1 or else contain a path of order $2c + 2$. In the latter case, we may observe that there exists a vertex v (which may be in either X or Y) with $d(v) \leqslant c$ or else every connected component of G is isomorphic to $K(c + 1, c + 1)$. To see this, note that if every vertex has degree $\geqslant c + 1$ then every vertex is an endvertex of a path of order $2c + 2$. Now the fact that there is no path of order $2c + 3$ together with the assumed degree condition first force this path to yield a cycle of order $2c + 2$ and then force these $2c + 2$ vertices to span a $K(c + 1, c + 1)$ component. Indeed, once G contains a cycle of order $2c + 2$, no further assumption need be made concerning degrees in order to conclude that G has a component isomorphic to $K(c + 1, c + 1)$. This follows from the fact that the $2c + 2$ vertices of the cycle must span a connected component of G which we may take to be [to maximize $q(G)$] $K(c + 1, c + 1)$.

Let us define $G_0 \subset K(n, n)$ using the scheme of Section 2 with $s = t = n - 1$, $M_{11} = M_{12} = M_{22} = 1$, and $M_{21} = 0$. Thus G_0 may be pictured as obtained by attaching a single pendant edge to $K(n - 1, n)$. The following result shows that G_0 is the unique extremal graph in a corollary to a theorem of Las Vergnas [5]. Berge presents both the theorem and corollary [2, Corollary 6, p. 216].

Lemma 4. If $|X| = |Y| = n$ and $G(X, Y, E)$ is not Hamiltonian, then $q(G) \leqslant n^2 - n + 1$. Furthermore, equality holds iff $G \simeq G_0$.

Proof. Suppose that $q(G) \geqslant n^2 - n + 1$. In Theorem 1 we found that $f_0(n, n, n - 1) = n^2 - n$. Therefore, we may assume that $(x_1, y_1, \ldots, x_m, y_n)$ is a path of order $2n$ in G. Let $S = \{j \mid x_1y_j \in E\}$ and let $T = \{j \mid x_jy_n \in E\}$. Since G is not Hamiltonian, it follows that $S \cap T = \emptyset$, and from this fact we may conclude that $q(G) \leq n^2 - n + 1$. By the same token, $q(G) = n^2 - n + 1$ means that (S, T) is a partition of $\{1, 2, \ldots, n\}$. With no loss of generality we may assume that $|S| \leq |T|$, so that $|S| \leq \frac{1}{2}n$. We claim that $S = \{1\}$. To see that this is the case, suppose to the contrary that $k \in S$ where $k \neq 1$. There is now a Hamiltonian path from x_k to y_n so x_k must behave in every respect as did x_1. In particular $\{j \mid x_k y_j \in E\} = S$. Now, taking into account the fact that both x_1 and x_k are adjacent to at most $\frac{1}{2}n$ vertices of Y, we find $q(G) \leq n^2 - 2(\frac{1}{2}n) = n^2 - n$, contrary to fact. Finally, since $S = \{1\}$, all edges missing in G are accounted for, so that $G \simeq G_0$ as claimed. ∎

The preceding lemma is used in the proof of the following special case of Theorem 2.

Lemma 5. The statement of Theorem 2 is true in case $a = b = c + 2$.

Proof. With $|X| = |Y| = c + 2$, let $G(X, Y, E)$ be a simple bipartite graph with $q(G) \geq c^2 + 2c + 2$ and suppose that G contains no path of order $2c + 3$. We wish to prove that the only possibilities are (i) $G \simeq G_{13}$, (ii) $G \simeq G_{14}$ (referring to Table II). The remark made near the beginning of this section may be used to note that if G contains a cycle of order $2c + 2$ then $G \simeq G_{13}$. In this case, G has $K(c + 1, c + 1)$ as a component and then its structure is completely determined. Henceforth, we shall assume that G contains no cycle of order $2c + 2$. Again, in view of an earlier remark, we may assume that $(x_0, y_0, \ldots, x_c, y_c)$ is a path of order $2c + 2$ in G. Let $H = G - \{x_0, y_c\}$. In view of the fact that G contains no cycle of order $2c + 2$, the following facts emerge: (1) $d(x_0) + d(y_c) \leq c + 1$, (2) $q(H) \leq c^2 + c + 1$. (The latter fact is a consequence of Lemma 4.) Now a count of edges and another application of Lemma 4 shows that (1) $d(x_0) + d(y_c) = c + 1$ and (2) $H \simeq G_0$. Now the $c + 1$ edges joining $\{x_0, y_c\}$ with $V(H)$ are completely locked in if a path of order $2c + 3$ is to be avoided and we obtain $G \simeq G_{14}$. ∎

Before presenting the proof of Theorem 2, it may be helpful if we make an observation concerning the general nature of the proof. The observant reader will have noted an alternative proof of Theorem 1. Instead of treating the connected and disconnected cases separately, we could have done both at once in an inductive argument. By deleting the appropriate vertices from G, we obtain an extremal graph H, the structure of which is then known precisely. Then it is a comparatively simple matter to check that the restoration of the deleted vertices and their incident edges can be done in only one way if the appropriate path is to be avoided. Thus the structure of G is determined. This is the tack which we shall follow in the proof of Theorem 2.

Proof of Theorem 2. The proof will be by induction on $a + b$. The reader is reminded that $c > 2$, $G \in F_1(a, b, c)$, and that $q(G) \geqslant f_1(a, b, c)$. We wish to prove that G is isomorphic to one of the graphs listed in Table II. The cases for us to consider are as follows.

$a \leqslant c$ or $a = b = c + 1$. This case is trivial since G must be complete bipartite.

$b > a = c + 1$ or $c + 1 < a < 2(c + 1)$. Let us first entertain the possibility that one component of G is isomorphic to $K(c + 1, c + 1)$. A simple calculation shows that with a and b as delimited above, the inequality $a + (b - 1)c > (c + 1)^2 + \{a - (c + 1)\}\{b - (c + 1)\}$ holds unless (i) $a = c + 1$, $b = c + 2$; (ii) $a = b = c + 2$; or (iii) $a = 2c + 1$. In each of these cases, equality holds. As a consequence, we find all cases in which G has $K(c + 1, c + 1)$ as a component. In particular, we find (i) $G \simeq G_{12}$, (ii) $G \simeq G_{13}$, or (iii) $G \simeq G_{15}$. Henceforth we assume that G has no $K(c + 1, c + 1)$ component. Recall that this means that for the path $(x_0, y_0, \ldots, x_c, y_c)$, which exists by virtue of Theorem 1, $d(x_0) + d(y_c) \leqslant c + 1$ and so $d(y_c) \leqslant c$. There are two subcases to consider.

$a = b$. The case of $a = b = c + 2$, which serves to anchor this induction, was dealt with in Lemma 5. Let x_0, y_c be as above and consider $H = G - \{x_0, y_c\}$. Using the induction hypothesis, we find that $H \in E_1(a - 1, a - 1, c)$ and that $d(x_0) + d(y_c) = c + 1$. Now it is easy to check that the restoration of x_0 and y_c yields $G \simeq G_{14}$.

$a < b$. Let y_c be as above and consider $H = G - \{y_c\}$. Using the induction hypothesis, we find that $H \in E_1(a, b - 1, c)$ and that $d(y_c) = c$. Restoration of y_c yields $G \simeq G_{14}$.

$a = 2c + 1$ or $a = b = 2(c + 1)$. As in the preceding case, the assumption that G has a component isomorphic to $K(c + 1, c + 1)$ gives $G \simeq G_{15}$ immediately. Otherwise, let y_c be as above and consider $H = G - \{y_c\}$. If $a = b = 2(c + 1)$, then H must fulfill impossible conditions. The case of $a = 2c + 1$ is considered above.

$b > a = 2(c + 1)$ or $a > 2(c + 1)$. A simple edge count shows that $G \not\simeq nK(c + 1, c + 1)$ with $n \geqslant 3$. As a consequence, there exists a vertex $y \in Y$ with $d(y) \leqslant c$. Consider $H = G - \{y\}$ and note that in all cases except $a = 2c + 2$, $b = 2c + 3$, H is an extremal graph and $d(v) = c$. In these cases, restoration of v yields $G \simeq G_{16}$. Now consider the case in which $a = 2c + 2$ and $b = 2c + 3$. First, let us suppose that there exists a vertex $x \in X$ with $d(x) \leqslant c - 1$. Then $H = G - \{x\}$ is extremal and $d(x) = c - 1$. Inspection of the possibilities for H shows that the restoration of x to produce $G \in F_1(2c + 2, 2c + 3, c)$ is impossible. Now suppose that there exists a vertex $y \in Y$ with $d(y) \leqslant c - 2$. Then $H = G - \{y\}$ is extremal and $d(y) = c - 2$. Since $c > 2$, we again find that restoration of y to produce $G \in F_1(2c + 2, 2c + 3, c)$ is impossible. Now by Theorem 1, either $G \simeq G_{04} \simeq G_{16}$ or else G contains a path $(x_0, y_0, \ldots, x_c, x_c)$ with $d(x_0) \geqslant c$ and $d(y_c) \geqslant c - 1$. Since $c > 2$, the latter case gives $d(x_0) + d(y_c) \geqslant c + 2$, and so, by observations made earlier, we find that G has a component isomorphic to $K(c + 1, c + 1)$. The deletion of this component must produce a graph $H \in F_1(c + 1, c + 2, c)$ with $q(H) \geqslant c^2 + 3c - 1$. Since $c > 2$, this is impossible. In summary, all roads lead to $G \simeq G_{16}$. This completes the proof of this case and so of the theorem. ∎

The final theorem is simply a "mopping up" of the case of paths of odd order. First, let us define two special graphs. The first, denoted G_{17}, is the disjoint union of $K(3, 3)$ and the graph G_{12} with parameters $a = 3$, $b = 4$, $c = 2$. The second, denoted G_{18}, is the disjoint union of $K(3, 3)$ and G_{14} with $a = 3$, $b = 4$, $c = 2$. Each of these newly defined graphs belongs to $F_1(6, 7, 2)$ and has 18 edges. In constructing extremal graphs for $E_1(a, b, 1)$, it is natural to look for components isomorphic to $K(2, 2)$. Another feasible component can be described by using the language of Section 2 with $s = t = 1$, $M_{11} = M_{12} = M_{21} = 1$, $M_{22} = 0$. In fact, all extremal graphs in $E_1(a, b, 1)$ have only the aforementioned components. The proof of the first part of the "mopping up" theorem just involves a careful review of the proof of Theorem 2 using $c = 2$. The second part is just a reflection of the remarks just made. Both proofs are left to the reader.

Theorem 3. Theorem 2 remains valid for $c = 2$ if Table II is supplemented by G_{17} and G_{18}. In the case of $c = 1$, the extremal function is given by

$$f_1(a, b, 1) = \begin{cases} a + b & \text{if } a = b = 2k, \\ a + b - 1 & \text{otherwise.} \end{cases}$$

References

[1] M. Behzad, G. Chartrand, and L. Lesniak-Foster, *Graphs and Digraphs*. Prindle, Weber and Schmidt, Boston (1979).

[2] C. Berge, *Graphs and Hypergraphs*. North-Holland, Amsterdam (1973).

[3] P. Erdös and T. Gallai, On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar.* 10 (1959) 337–356.

[4] B. Jackson, Cycles in bipartite graphs. *J. Combinatorial Theory Ser. B* 30 (1981) 332–342.

[5] M. Las Vergnas, Sur l'existence des cycles Hamiltoniens dans un graphe. *C. R. Acad. Sc.* 270 (1970) A-1361–1364.