The Solution to an Extremal Problem on Balanced Extensions of Graphs*

A. Ruciński

DEPARTMENT OF DISCRETE MATHEMATICS ADAM MICKIEWICZ UNIVERSITY POZNAŃ, POLAND

A. Vince

DEPARTMENT OF MATHEMATICS UNIVERSITY OF FLORIDA GAINESVILLE, FLORIDA, USA

ABSTRACT

For *n* sufficiently large the order of a smallest balanced extension of a graph of order *n* is, in the worst case, $[(n + 3)^2/8]$. © 1993 John Wiley & Sons, Inc.

1. INTRODUCTION

Graphs in this paper will be finite, undirected, without loops and multiple edges. Throughout v(G) = |V(G)| and e(G) = |E(G)|. Let d(G) denote the density of graph G

$$d(G)\frac{e(G)}{v(G)}.$$

Call G balanced if $d(H) \le d(G)$ for all subgraphs H of G. Balanced graphs originated in the work of Erdös and Rényi on random graphs [3], in a result giving the probability that a random graph contains a given graph

^{*}This paper was written during the first author's visit to Department of Mathematics, University of Florida, Gainesville, in the fall of 1990.

Journal of Graph Theory, Vol. 17, No. 3, 417–431 (1993) © 1993 John Wiley & Sons, Inc. CCC 0364-9024/93/030417-15

G. Subsequently, other applications of balanced graphs to random graph theory have appeared in papers by Bollobás [2], Karoński and Ruciński [5], Ruciński and Vince [8], and Spencer [10]. If G is not balanced, then G contains a subgraph with greater density than G. In particular, let m(G) be the maximum density of a subgraph of G

$$m(G) = \max_{H\subseteq G} d(H).$$

Note that G is balanced if and only if m(G) = d(G). Throughout the paper, m(G) will be called the *global density* of G.

Any graph G can be embedded as a subgraph of a balanced graph F; for example, take F to be the complete graph on the vertices of G. It is not so obvious that F can be chosen so as not to increase the global density. A graph F is called a *balanced extension* of G if

(1)
$$G \subseteq F$$
,
(2) F is balanced, and
(3) $d(F) = m(G)$.

It was proved by Győri, Rothschild, and Ruciński [4], and independently by Payan [7], that every graph has a balanced extension. The first proof of the existence of balanced extensions was simplified by Ruciński and Vince [8] by introducing strongly balanced graphs, and in [9], inspired by a question of Erdös, the authors go on to ask for the minimum balanced extension of a graph. More precisely, let ext(G) denote the minimum number of vertices in any balanced extension of G

$$ext(G) = min\{v(F): F \text{ is a balanced extension of } G\}.$$

An interesting interpretation of ext(G) was pointed out in [7]. The bicircular matroid of a graph is the matroid whose independent sets are the edge-sets of the subgraphs with at most one cycle in each component. Edmonds' matroidal generalization of the Nash-Williams arboricity theorem [1, p. 489, cor. 1], when applied to the bicircular matroid, states that the edges of G can be covered by [m(G)] subgraphs, each with at most one cycle in each component. Thus, if d(G) = m(G) = [m(G)], then the edges of G can be covered by d(G) edge-disjoint subgraphs, each with exactly one cycle in each component. Let m(G) = s/t. Upon replacing each edge of G by t parallel edges, we arrive at the conclusion that ext(G) is the smallest number of vertices in a supergraph of G that can be covered by s subgraphs, each with exactly one cycle in each component, in such a way that each edge of G belongs to exactly t of the subgraphs.

Let *n* always denote the order of a graph *G*. The constructive proof in [8] gives the bound $ext(G) < (1 + o(1))(n^2/4)$. Erdös conjectured that for sufficiently dense graphs the bounds on ext(G) can be improved, in fact that ext(G) is linear in *n*. In this direction Ruciński and Vince [9] proved that

if $e(G) > cn^2$, $0 < c < \frac{1}{2}$, then ext(G) < c'n where c' depends only on c. Łuczak and Ruciński [6] used nonconstructive random graph techniques to extend this result to less dense graphs. They showed that for some, quite large, constant c any graph G with sufficiently many vertices satisfies

$$ext(G) < \begin{cases} \frac{cn}{m(G) - 1}, & \text{if } 1 < m(G) < 1\frac{1}{9}, \\ 203n, & \text{if } m(G) > 4\frac{1}{4}. \end{cases}$$

If $e(G) > cn^2$ for some constant c then m(G) > cn. Therefore, the second part of the [6] result implies the result of Ruciński and Vince from [9]. If the global density of G is small, however, say of order m(G) = 1 + (c'/n), the first part of the result only yields a $(c/c')n^2$ bound. The problem of proving an appropriate bound in the disturbing gap between $1\frac{1}{9}$ and $4\frac{1}{4}$ remains open. If $m(G) \le 1$ then all connected components of G are trees or unicyclic graphs. In this case it is easy to show that $ext(G) \le 2n$. As soon as m(G) > 1 the situation changes; there are graphs with ext(G) as large as $n^2/8$.

The aim of this paper is to find ext(G) in the worst case. Let us, therefore, define the extension number

$$\operatorname{ext}(n) = \max_{\nu(G)=n} \operatorname{ext}(G).$$

The results described above give the bounds $n^2/8 < ext(n) < (1 + o(1)) \times (n^2/4)$, which appeared in [9]. As a consequence of the result proved in the present paper, it turns out that asymptotically $n^2/8$ is the correct order of ext(n). Moreover, our main result gives an exact value of ext(n) for n sufficiently large.

Theorem. For *n* sufficiently large

$$\operatorname{ext}(n) = \left\lfloor \frac{(n+3)^2}{8} \right\rfloor.$$

In the existence proof in [8], the balanced extension F of G can be constructed so that G is an *induced* subgraph of F. Likewise, our theorem is strong in the sense that, for n sufficiently large, the following holds. Every graph G has a balanced extension F with order at most $\lfloor (n + 3)^2/8 \rfloor$ and such that G is an *induced* subgraph of F. On the other hand, there exists a graph G whose any balanced extension F has order at least $\lfloor (n + 3)^2/8 \rfloor$, whether or not G is required to be an induced subgraph of F.

Because the proof of the theorem is somewhat involved, the lower bound, as well as a sketch of the upper bound, are contained in Section 2. This is intended as motivation for the main Lemma 5 in Section 3 and to give insight into the detailed proof of the Theorem in Section 4.

2. THE LOWER BOUND AND A SKETCH OF THE UPPER BOUND

Proof of the Lower Bound. To show the lower bound $ext(n) \ge \lfloor (n+3)^2/8 \rfloor$, it is sufficient to construct, for sufficiently large n, a graph G_n such that $v(F) \ge \lfloor (n+3)^2/8 \rfloor$ for any balanced extension F of G_n . This is done as follows. For every $n \ge 7$, let $n = n_1 + n_2$ be the unique partition of n into the sum of two integers such that $1 \le n_2 - n_1 \le 4$ and n_1 is odd. (For $n \equiv 0, 1, 2, 3 \pmod{4}$ the value of n_2 is, respectively, (n/2) + 1, (n+3)/2, (n/2) + 2, (n+1)/2.) Let graph G_n consist of an n_2 -cycle, a chord, and n_1 pendant vertices adjacent to the same vertex of the cycle (G_{11} appears in Figure 1a). We will show that

$$ext(G_n) \ge \frac{1}{8}n^2 + \frac{3}{4}n + \frac{5}{8}$$

This will give the lower bound $ext(n) \ge \lfloor (n+3)^2/8 \rfloor$, since $\frac{1}{8}n^2 + \frac{3}{4}n + \frac{5}{8}$ differs by only $\frac{1}{2}$ from $(n+3)^2/8$.

Assume F is a balanced extension of G_n . Since $d(F) = m(G_n) = 1 + (1/n_2) > 1$, it is easy to show that every vertex of F has degree at least 2. Obviously, for each vertex w of G_n , $\deg_F(w) \ge \deg_{G_n}(w)$. Moreover, since



 n_1 is odd, either this inequality is strict for at least one of the n_2 vertices of the cycle or deg_F(w) ≥ 3 for at least one of the remaining $v(F) - n_2$ vertices of F. Hence,

$$2m(G_n)\nu(F) = 2e(F) = \sum_{w \in V(F)} \deg_F(w)$$

$$\geq \sum_{w \in V(G_n)} \deg_{G_n}(w) + n_1 + 2(\nu(F) - n) + 1$$

$$= (n_1 + 2) + 6 + 2(\nu(F) - 3) + 1$$

$$= n_1 + 3 + 2\nu(F),$$

which is equivalent to

$$v(F) \ge \frac{n_1+3}{2(m(G_n)-1)} = \frac{1}{2}n_2(n_1+3) \ge \frac{1}{8}n^2 + \frac{3}{4}n + \frac{5}{8},$$

which, in turn, implies $v(F) \ge \lfloor (n + 3)^2/8 \rfloor$.

Because the proof of the upper bound $ext(n) \le |(n + 3)^2/8|$ is more complicated than that of the lower bound, a sketch of the proof is given here. It must be shown that any graph G with n vertices (sufficiently large) has a balanced extension F with $v(F) \leq \lfloor (n+3)^2/8 \rfloor$. The following definition is required. For any graph G, let G^o , called the balanced core of G, be the largest subgraph in G that realizes the global density m(G). In fact, the balanced core is the union of all subgraphs of G that achieve its global density, i.e., $G^o = \bigcup \{H: d(H) = m(G)\}$. Note that a graph G is balanced if and only if $G^o = G$. Figure 1b shows a balanced extension F of G_{11} that achieves the above lower bound so that $ext(G_{11}) = 24$. With the graph G_{11} as an example, the idea of the proof of the upper bound is as follows. The balanced extension F of a graph G is built in stages $G = F_0, F_1, F_2, \dots, F_t = F$. At each stage (the last stage is a special case) a graph B_i is adjoined to the previous graph F_{i-1} at exactly two vertices not in F_{i-1}^{o} to form the new graph F_i . In the example, B_1 and B_2 are paths on 6 vertices. At the last stage B_3 is a cycle on 6 vertices. In the example $m(G_{11}) = 7/6$ and at each stage the balanced core F_i^o of F_i consists of the subgraph induced by all nonpendant vertices. In general, at each successive stage, F_i^o includes the previous balanced core F_{i-1}^o and the adjointed graph B_i . At each stage (except possibly the last and except the special case when G^{o} is a complete graph) at least two additional vertices of the original G are included in the balanced core. When all vertices of the original G are included, the procedure terminates. So at the last stage $F^o = F$ and hence F is balanced. At each stage the adjoined graph B_i must be, in a certain sense, extremely balanced to ensure that the balanced property is not destroyed and also must be small to ensure the desired upper bound. The construction of such graphs is the purpose of Section 3 and Lemma 5 in particular. If the

graphs B_i are adjoined to just one vertex of G instead of two, as in [8], the number of stages, in the worst case, doubles and the procedure results in only an $n^2/4$ upper bound for ext(n).

Comment. The only known graphs G that achieve the maximum extension number are those whose balanced core satisfies $e(G^o) = v(G^o) + 1$. On the other hand, the best lower bound we can get on ext(G) for graphs with at least two more edges than vertices in their balanced core is of the order of $n^2/12$, a jump from $n^2/8$. In fact, for every $k \ge 2$ and n sufficiently large there is a graph $G_{n,k}$ whose core has k more edges than vertices and for which

$$\operatorname{ext}(G_{n,k}) \geq \frac{n^2}{4(k+1)}.$$

It seems that by a suitable refinement of the methods in this paper one can prove that this is indeed the correct asymptotic value of ext(G) in the worst case. This would mean that in the asymptotic behavior of ext(n) jumps do occur.

3. DOUBLE ROOTED BALANCED GRAPHS

This section primarily concerns the construction of highly balanced graphs B(v, e), with v vertices and e edges, and with two roots b_1 and b_2 , needed in the proof of the main theorem. It will be assumed throughout this section that v > 4. Define r equidistributed points (modulo s) on the set $\{1, 2, \ldots, s\}$ as follows:

$$R = R(s,r) = \left\{ \left\lceil \frac{s}{r} \right\rceil, \left\lceil \frac{2s}{r} \right\rceil, \ldots, \left\lceil \frac{rs}{r} \right\rceil \right\} = \left\{ u_1, u_2, \ldots, u_r \right\}.$$

Denote by C(s,t) the graph with vertex set $\{1,2,\ldots,s\}$ and edge set $\{\{i,j\}: |i-j| \le t\}$, where the subtraction is modulo s. Denote by D(s,t) the graph obtained by joining two new vertices b_1 and b_2 to all vertices of C(s,t). Denote by A(s,t) the graph obtained by joining a new vertex b_1 to all the even elements of C(s,t) and a new vertex b_2 to all the odd elements of C(s,t). The graphs B = B(v,e) and their roots are now defined by cases. In several of these cases, integers s, k, r are defined by s = v - 2 and e = sk + r, where $0 \le r < s$.

Case 1. e = v - 1. Here B is a path. The two end points of B are the roots b_1 and b_2 .

Case 2. e = v, v even. Here B is a cycle. Two diametrically opposite vertices of the cycle are the roots b_1 and b_2 .

Case 3. e = v + 1. Here B consists of a cycle $\{1, 2, ..., v\}$ with a chord joining vertices 1 and $\lfloor v/2 \rfloor$. The vertices $\lfloor v/4 \rfloor$ and $\lfloor 3v/4 \rfloor$ are the roots b_1 and b_2 .

Case 4. $v + 2 \le e < 2v - 4$ (or e = s + r, $4 \le r < s$). Here B is obtained from the cycle C(s, 1) by adding two extra vertices b_1 and b_2 of equal degree $\lfloor r/2 \rfloor$ joined to the elements of R in an alternating fashion. In addition, if r is odd, one long chord from $u_r = s$ to $\lfloor s/2 \rfloor$ is added. More precisely, $E(B) = E(C(s, 1)) \cup \{\{b_i, u_j\}: i \equiv j \pmod{2}, i = 1, 2, j = 1, 2, ..., 2\lfloor r/2 \rfloor\} \cup \{s, \lfloor s/2 \rfloor\}$, the last union only if r is odd.

Case 5. $2v - 4 \le e < 3v - 6$, v > 6 (or e = 2s + r, $0 \le r < s$, s > 4). Here B is obtained from A(s, 1) by adding the set of r edges $L = \{\{i, i + 2\}, i \in R\}$ where addition is modulo s.

Case 6.

$$3v - 6 \le e \le \begin{cases} \binom{v}{2} - 1, & \text{if } v \text{ is odd,} \\ \frac{v(v - 2)}{2}, & \text{if } v \text{ is even,} \end{cases}$$

(or e = ks + r, $0 \le r < s$, $3 \le k \le \lfloor (s + 3)/2 \rfloor$). Here B is obtained from D(s, k - 2) by adding the set of r edges $L = \{\{i, i + k - 1\}, i \in R\}$ where addition is modulo s.

Case 7. $(v(v-2))/2 < e \le {v \choose 2} - 1$, v even. Here B is obtained from the complete graph K_{v-1} by adding an extra vertex b_1 and any $e - {v-1 \choose 2}$ edges from b_1 to K_{v-1} . The other root b_2 is any vertex in K_{v-1} not adjacent to b_1 .

Note that every pair (v, e), $v - 1 \le e \le {\binom{v}{2}} - 1$, falls into one of the seven cases except (5, 7), (5, 8), (6, 8), (6, 9), (6, 10), (6, 11), and $\{(v, e): e = v, v \text{ odd}\}$. The latter set of pairs is treated as a special case in the proof of the main theorem. Note also that in all cases the roots b_1 and b_2 are nonadjacent. The crucial properties of the graphs B(v, e) are stated in Lemma 5 below. Letting v = v(G), e = e(G), the following notation will be used in the lemmas. For i = 0, 1, or 2 and i < v,

$$d_i(G)=\frac{e}{v-i},$$

whereas if v = i define $d_i(G) = 0$. Note that $d_0 = d$ is the ordinary density. Also, for a given graph G the *deficit functions* of graph H with respect to G are defined by

$$f_i(H) = d_i(G)(v(H) - i) - e(H).$$

The degree of vertex u in graph G will be designated by $\deg_G(u)$. The proofs of Lemmas 1-4 follow immediately from the definitions and the first three lemmas are similar to results that appear in [8] and [9]. The exception in Lemma 1 (because $v(K_2) = 2 = i$) will never occur hereafter.

Lemma 1. $f_i(H) \ge 0$ if and only if $d_i(H) \le d_i(G)$, with the single exception i = 2, $G = H = K_2$.

Lemma 2. Let u be any vertex of a graph H with at least three vertices. Then $f_i(H) \le f_i(H - u)$ if and only if $d_i(G) \le \deg_H(u)$.

Lemma 3. For graphs H_1 and H_2

$$f_i(H_1 \cup H_2) = f_i(H_1) + f_i(H_2) - f_i(H_1 \cap H_2).$$

Lemma 4. A set of x consecutive elements from the set $\{1, 2, ..., s\}$ modulo s contains less than xr/s + 1 elements from R(s, r).

Lemma 5. For every pair of integers (v, e), $v - 1 \le e \le {v \choose 2} - 1$, with $v \ge 10$ and with the exception of $\{(v, e): e = v, v \text{ odd}\}$, there exists a graph B = B(v, e) containing two distinguished nonadjacent vertices b_1 and b_2 with the following properties:

- (1) $d(H) \le d(B)$ for all subgraphs H of B.
- (2) $d_1(H) \le d_1(B)$ for all subgraphs H containing at least one of the vertices b_1 or b_2 .
- (3) $d_2(H) \le d_2(B)$ for all subgraphs *H* containing both vertices b_1 and b_2 .

Remark. If e = v, v odd, we take for B a cycle of length v - 1 with a pendant edge. The vertex of degree one is then b_1 and the vertex diametrically opposite the neighbor of b_1 on the cycle is b_2 . It is easy to check that this graph satisfies conditions (1) and (3) of Lemma 5, but not condition (2). This graph will be used in the proof of the main theorem.

Proof. The following principles are used in the proof. Let f_i , i = 0, 1, 2, be the deficit functions with respect to B. In checking each of the seven cases there is, due to Lemma 1, the freedom to prove, for all the relevant subgraphs H of B, either that $f_i(H) \ge 0$ or that $d_i(H) \le d_i(B)$, whichever is more convenient.

It is easy to check that if condition (3) holds for a particular subgraph H, then condition (2) also holds for H; likewise, if condition (2) holds for H, then so does condition (1). Thus if H contains both vertices

 b_1 and b_2 , only condition (3) is checked; if H contains one of the vertices b_1 or b_2 , then only condition (2) is checked; and if H contains neither of the vertices b_1 and b_2 , then condition (1) is checked. Also in checking condition (1) it may be assumed that H is connected. To see this, suppose $H = H_1 \cup H_2$, where H_1 and H_2 are disjoint. Then by Lemma 3 we have $f_0(H_1 \cup H_2) = f_0(H_1) + f_0(H_2)$. If the left-hand side is negative, then at least one of the terms on the right-hand side must also be negative. Similarly, concerning condition (2) it may be assumed that H is not the union of two graphs that intersect at exactly one of the vertices b_1 or b_2 ; and concerning condition (3) it may be assumed that H is not the union of two graphs that intersect exactly at vertices b_1 and b_2 . The above facts imply that $C = H - \{b_1, b_2\}$ can always be assumed connected. In this proof x always denotes the number of vertices in C.

Moreover, it is sufficient to prove conditions (1)-(3) only for a subgraph H of B that minimizes the deficit function f_i . It will always be assumed that H is such a subgraph. By this special choice of H and Lemma 2, no vertex u of H has $\deg_H(u) < d_i(B)$ and no vertex not in H has more than $d_i(B)$ neighbors in H. In particular, H contains no pendant vertex as long as $d_i(B) > 1$.

Now each of the three conditions in the lemma must be checked for each of the cases of B(v, e) defined above. The cases (1)–(3) and (7) are routine to check. In checking case (4), let the x vertices on C(s, 1) form t arcs of consecutive vertices. Since C is connected, t = 1 or 2. Let \overline{r} be the number of elements of R among these x vertices. By Lemma 4, $\overline{r} \leq (rx/s) + t$. To check condition (3) in the lemma $d_2(H) \le (x - t + \overline{r})/x \le 1 + (r/s) =$ $d_2(B)$. If both b_1 and b_2 are in H, then condition (2) follows from condition (3). Otherwise assume that, say, b_1 is in H and that b_2 is not. Then, by Lemma 2, there is at most one element of $\{u_1, \ldots, u_{2|r/2|}\}$ with an even index among the x vertices on C(1, s), because otherwise b_2 has at least $2 > d_1(B)$ neighbors in H. This restricts H to a few possible forms. Checking them case by case confirms that e(H) = x + 1 or x + 2. Moreover, if e(H) =x + 1, then $x \ge (3s/2r)$, whereas if e(H) = x + 2, then $x \ge (3s/r)$. In either case $d_1(H) \le 1 + (2r/3s) \le 1 + ((r-1)/(s+1)) = d_1(B)$ if $s \ge 8$ (i.e., $v \ge 10$). If at least one of b_1 or b_2 is in H, then condition (1) follows from condition (2). Otherwise $d(H) \le 1 + (1/s) < d(B)$.

To prove case (5), first consider condition (3). Since C is connected, either consecutive vertices among these x are spaced one apart or two apart. The second possibility can be eliminated by Lemma 2 because the intermediate vertex would have at least $3 > d_2(B)$ neighbors in H. Using Lemma 4 we have $d_2(H) \le (1/x)(2x - 1 + (xr/s) + 1) = 2 + (r/s) =$ $d_2(B)$. If both b_1 and b_2 are in H, then condition (2) follows from condition (3). Otherwise assume that b_1 is in H and that b_2 is not in H. Using Lemma 4 we have $d_1(H) \le (1/x)x - 1 + (x + 1)/2 + (rx/s) +$ $1) = (3/2) + (1/2x) + (r/s) \le 2 + (r - 2)/(s + 1) = d_1(B)$ for $s \ge 8$ (i.e., $v \ge 10$) and $x \ge 3$. (For x = 2 we have $d_1(H) \le (3/2) < d_1(B)$.) To check condition (1) it again may be assumed that neither b_1 nor b_2 are in H, and that H is connected. Then $d_0(H) \le (1/x)(x + (rx/s)) \le d_0(B)$ for $s \ge 4$ (i.e., $v \ge 6$).

In case (6) we shall prove that *all* subgraphs of *B*, and not only those containing b_1 and b_2 , satisfy (3). This, by the remarks above, will imply conditions (1) and (2). Let \overline{r} denote the number of edges of *R* contained in *H*. The proof proceeds by induction on *k*. First consider k = 3. If $x \ge 4$, then by Lemma 2 vertices b_1 and b_2 are in *H* since $4 \ge d_2(B)$. The cases x < 4 can be routinely checked. As in case (5), consecutive vertices among the *x* vertices are spaced one apart. Now $d_2(H) \le (1/x)(3x - 1 + \overline{r}) \le 3 + (r/s) = d_2(B)$. For k > 3, assume that condition (3) holds for *all* subgraphs *H* of B(s + 2, sk + r), i.e., $f_2(H) \ge 0$ for all subgraphs *H*. Let *H'* be any subgraph of B(s + 2, s(k + 1) + r) and let *H* be the subgraph of B(s + 2, sk + r) induced by the vertices of *H'*. Then $e(H') \le e(H) + x$ implies $f_2(H') - f_2(H) = [(k + 1 + (r/s))x - e(H')] - [(k + (r/s))x - e(H)] \ge x - x = 0$. Therefore $f_2(H') \ge f_2(H) \ge 0$.

4. MAIN RESULT

This section is entirely devoted to a detailed proof of the upper bound in the main theorem. The lower bound was proved in Section 2.

Theorem. For *n* sufficiently large

$$\operatorname{ext}(n) = \left\lfloor \frac{(n+3)^2}{8} \right\rfloor.$$

Proof of the Upper Bound. To obtain the bound $ext(n) \le (n + 3)^2/8$, let G be an n-vertex graph. We may assume that 1 < m = m(G) < 4.25, since outside this range ext(G) is linear in n (cf. Section 1). A balanced extension F of G will be constructed in stages so that $v(F) \le (n + 3)^2/8$. Set $n^o = v(G^o)$ and express m = p/q, where p and q are relatively prime integers. A sequence of graphs

$$G = F_0 \subset F_1 \subset \cdots \subset F_t = F$$

will be constructed such that F is a balanced extension of G and for i = 1, ..., t.

(1) F_{i-1} is an induced subgraph of F_i ; (2) $m(F_i) = m$; (3) setting $n_i = v(F_i)$, $n_i^o = v(F_i^o)$, $\overline{n}_i = n_i - n_i^o$, $\Delta_i = \overline{n}_{i-1} - \overline{n}_i$, i = 0, 1, ..., t; (a) $n_i - n_{i-1} \le q + 8$ and $\Delta_i \ge \max(2, q - 7) - 1$, (b) $n_i - n_{i-1} \le n^o - 2$ and $\Delta_i \ge 2$, or (c) $n_i - n_{i-1} \le n^o - 1$ and $\Delta_i \ge 1$ —option (c) possible only for i = t - 1 or if G^o is a complete graph.

Moreover, both bounds in (a) can be improved by 1 except when i = t - 1. Note that $\overline{n}_0 = n - n^o$, and by (3), \overline{n}_i is reduced at each stage by at least 1. At the first t such that $\overline{n}_t = 0$ we have $F_t^o = F_t$ that, by the comments in Section 2 and assuming condition (2), implies that F_t is the desired balanced extension of G. It follows from (3) that, for n large enough, $v(F) \leq (n + 3)^2/8$. To see this first note that

$$v(F) \leq n + \sum_{i=1}^{l} (n_i - n_{i-1}).$$

Four cases will be considered. Note that q divides n^o , so that, unless $q = n^o$ we have $q \le n^o/2$.

- 1. If $\max(q + 8, n^{o} 1) \le 17$ (it is easy to check that this includes the case when \overline{G} is complete), then $t \le n$ and $v(F) \le n + 17n = 18n$.
- 2. If $q \le n^{o}/2$ and $\max(q + 8, n^{o} 1) > 17$, then either (a) $q \ge 10$ and $q + 8 \le 2q - 2 \le n^{o} - 2$, (b) $n^{o} \ge 20$ and $q + 8 \le (n^{o}/2) + 8 \le n^{o} - 2$, (c) $n^{o} = 19, q = 1$ and $q + 8 \le 9 \le n^{o} - 2$. In any case, $q + 8 \le n^{o} - 2$ and $t \le \frac{1}{2}(n - n^{o} + 1)$, and consequently,

$$v(F) \leq n + \frac{1}{2}(n - n^{o} + 1)(n^{o} - 2) + 1 \leq \frac{(n + 3)^{2}}{8},$$

the last inequality obtained by maximizing with respect to n^o . The extra "1" here and in Case 4 below comes from the possibility that $n_{t-1} - n_{t-2} = n^o - 1$ and not $n^o - 2$.

3. If $q = n^o$ and (3a) holds for at least one *i*, then $t \le 1 + \frac{1}{2}(n - n^o - (q - 7) + 1)$ and

$$v(F) \le n + \frac{1}{2}(n - n^{o} - (q - 7) + 3)(q + 8) \le \frac{1}{16}n^{2} + O(n).$$

4. Finally, if $q = n^{\circ}$ and (3a) never occurs, we obtain the critical bound again:

$$v(F) \le n + \frac{1}{2}(n - n^o + 1)(n^o - 2) + 1 \le \frac{(n + 3)^2}{8}$$

It now suffices to describe the construction of the graphs F_i 's, and verify conditions (1)-(3). The first F_i 's serve to remove the isolated vertices of G if there are at least two of them. (Clearly, G^o has no isolated vertices.) Assume that $i_1, \ldots, i_s, s \ge 2$, are the isolated vertices of G. Set $s_0 = s$ if G^o is complete and $s_0 = \lfloor s/2 \rfloor$ otherwise. We build F_j , $j = 1, \ldots, s_0$, by adjoining to F_{j-1} a copy of G^o in such a way that it contains i_j and $n^o - 1$ new vertices if G^o is complete and it contains nonadjacent i_{2j-1}, i_{2j} , and $n^o - 2$ new vertices otherwise. Clearly, F_{j-1} is an induced subgraph of F_j . It is routine to check that the new balanced core F_j^o is a disjoint union of j copies of G^o , so that condition (2) holds. Also $n_j - n_{j-1} = n^o - 2$ or $n^o - 1$, and $\Delta_j = 2$ or 1, respectively, verifying condition (3b) or (3c). Thus, for $i \ge s_0$, there will be at most one isolated vertex in F_i .

Next the transition from F_i to F_{i+1} , for $s_0 \le i \le t - 1$, will be described. In order to unify our description, add an extra pendant vertex to F_{t-1} , if necessary, to ensure that $\overline{n}_{t-1} \ge 2$ (this adds a "1" to the bounds in (3a) and (3c) for i = t - 1). Hence, we may assume that $\overline{n}_i \ge 2$ for all $i \ge s_0$. The relevant deficit function to use in the proof is f(H) = mv(H) - e(H), i.e., the deficit function f_0 with respect to G^o already defined in Section 2. In particular, both Lemmas 1 and 3 remain valid for f. Notice that $m(F_i) = m$ if and only if both $f(H) \ge 0$ for all subgraphs H of F_i and $f(H_0) = 0$ for some subgraph H_0 of F_i . Let

$$f^* = \min\{f(H): H \subseteq F_i, |V(H) \setminus V(F_i^o)| \ge 2\}.$$

Applying the modularity of f one can check that the minimum is always achieved by a subgraph containing the balanced core F_i^o , i.e.,

$$f^* = \min\{f(H): F_i^o \subset H \subseteq F_i, \ v(H) - v(F_i^o) \ge 2\}.$$

Note that

 $(4) \ 0 < f^* \le 2m - 1,$

where the upper bound follows by taking F_i^o plus two vertices and one edge for *H*. This is always possible because there are at least two vertices outside the balanced core and at most one isolated vertex among them. Consider the equation in variables v and e:

(5)
$$mv - e = 2m - f^*$$
.

For any solution (v, e) of (5) it follows from (4) that

(6)
$$m(v-2) < e \leq mv - 1$$
.

Since m = p/q we can express $f^* = r/q$, where p, q, and r are positive integers. On clearing fractions, (5) becomes a linear Diophantine equation, and because p and q are relatively prime, it has infinitely many integer

solutions of the form $(v_0 + sq, e_0 + sp)$, where (v_0, e_0) is any particular solution. Hence there is always a solution (v, e) satisfying

(7)
$$10 \le v \le q + 9$$
.

In addition,

$$(8) \quad v-1 \leq e \leq \binom{v}{2} - 1$$

follows from (6), (7), and the fact that 2m + 1 < 10 (recall that m < 4.25). Let F^* be a subgraph of F_i realizing the minimum f^* , i.e., $F_i^o \subset F^* \subseteq$ $F_i, v^* = v(F^*) - v(F_i^o) \ge 2$, and $f(F^*) = f^*$. Let $x, y \in V(F^*) \setminus V(F_i^o)$. Let B be the graph B(v, e) in Lemma 5 of Section 2 unless v = e, v odd, in which case let B be the graph described in the remark following Lemma 5. (Inequality (8) allows us to apply Lemma 5.) We construct F_{i+1} by attaching a copy of B to F_i in such a way that b_1 and b_2 are identified with x and y, and that F_i and B are otherwise disjoint. Note that G is an induced subgraph of F_{i+1} at each stage. In the case when v = e is odd, if x is the only vertex of $V(F^*) \setminus V(F_i^o)$ joined by an edge to F_i^o , we must identify x with b_1 and not with b_2 . (We will need this special requirement at the very end of the proof.) In either case, by Lemma 3 and by (5) we have $f(F^* \cup B) = f^* + mv - e - 2m = 0$, and so, provided (2) is true, $F^* \cup B \subseteq F_{i+1}^o$. This means that $\Delta_i \ge v^*$. Thus to verify condition (3) we must show that $v^* \ge \max(2, q - 7)$. Recall that by the definition of F^* , $v^* \ge 2$. Also it follows from (5) and the fact that p and q are relatively prime that, for any solution (v', e') of (5), $v' + v^* - 2$ is a multiple of q. Hence $v' + v^* - 2 \ge q$ if v' > 0. If v, the solution satisfying (7), is not greater than q, then (3b) (or (3c)) holds. Otherwise, v' = v - q is another solution of (5), $0 < v' \le 9$, and therefore $v^* \ge q - v' + 2 \ge q - 7$. To complete the proof of (3) observe that $n_i - n_{i-1} = v - 2 \le q + 7$ by inequality (7). Now it only remains to check condition (2) $m(F_{i+1}) = m$, i.e., that for every subgraph H of F_{i+1} we have $f(H) \ge 0$. We assume inductively that $m(F_i) = m$. Thus there is nothing to check if $H \subset F_i$. So assume that at least one vertex of H is not in F_i . There is no loss of generality in assuming that $H \supset F_i^o$ because if f(H) < 0 then $f(H \cup F_i^o) = f(H) + f(F_i^o) - f(H \cap F_i^o) < 0$. Let $H_R = H \cap B$, $H_F = H \cap F_i$, $v_0 = v(H_B)$, and $e_0 = e(H_B)$. Recall that v = v(B) and e = e(B). Clearly, $V(H_B) \cap V(H_F) = V(H) \cap \{x, y\}$ and $H_B \cap H_F$ contains no edge. Consider 3 cases.

Case 1. $V(H) \cap \{x, y\} = \emptyset$.

In this case $f(H) = f(H_F \cup H_B) = f(H_F) + f(H_B)$. Since $H_F \subseteq F_i$, we know that $f(H_F) \ge 0$. By (6) and property (1) of Lemma 5, we have $\frac{e_0}{v_0} \le \frac{e}{v} < m$; therefore $f(H) \ge f(H_B) \ge 0$.

Case 2. $V(H) \supseteq \{x, y\}$.

Now, $f(H) = f(H_F \cup H_B) = f(H_F) + f(H_B) - 2m$ and $f(H_F) \ge f^*$ since $|V(H_F) \setminus V(F_i^o)| \ge 2$. By (6) and property (3) of Lemma 5,

$$\frac{e_0}{v_0-2} \leq \frac{e}{v-2} \Longleftrightarrow \frac{e-e_0}{v-v_0} \geq \frac{e}{v-2} > m.$$

The last inequality implies $f(H_B) > f(B)$, so $f(H) > f^* + mv - e - 2m = 0$ by (5).

Case 3.
$$V(H) \cap \{x, y\} = \{x\}$$
, say

Assume first that $e \neq v$ or v is even. In this case refine (4) to

(9) $0 < f^* \leq f(H_F) + m$,

by taking, if $f(H_F) < f^*$, H_F plus a vertex in the definition of f^* . (Note that $f(H_F) < f^*$ implies that $V(H_F) \setminus V(F_i^o) = \{x\}$.) By (5) and (9) combined and by condition (2) of Lemma 5,

$$\frac{e_0}{v_0-1} \leq \frac{e}{v-1} \leq m + \frac{f(H_F)}{v-1} \leq m + \frac{f(H_F)}{v_0-1},$$

which gives $f(H) = f(H_F \cup H_B) = f(H_F) + f(H_B) - m \ge 0$.

Turning to the special case v = e, v odd, where we cannot apply condition (2) of Lemma 5, note that (5) now becomes

(10) $f^* + v(m-1) - 2m = 0$.

Also f(B) = v(m-1), $e_0 \le v_0$, and $e_0 = v_0$ implies $v_0 \ge v - 1$. If $e_0 < v_0$, then

$$f(H) = f(H_F) + f(H_B) - m \ge v_0(m-1) + 1 - m > 0$$

since $v_0 \ge 2$ and m > 1. If $e_0 = v_0 = v$, i.e., if $H_B = B$, then, by (9) and (10),

$$f(H) = f(H_F) + f(H_B) - m = f(H_F) - f^* + m \ge 0.$$

Thus, it only remains to check the case $e_0 = v_0 = v - 1$. Since $f(H_B) > 0$ and $f(H_F) \ge 0$, we are done if either $f(H_F) \ge m$ or $f(H_B) \ge m$. If $m \ge 9/8$, then $f(H_B) \ge m$ for $v \ge 10$, which is the case by inequality (7). Otherwise, i.e., when $f(H_F) < m < 9/8$, we claim that (9) can be further improved to yield $f^* \le f(H_F) + m - 1$. In this case, using (5),

$$f(H) = f(H_F) + (v - 1)(m - 1) - m \ge (f^* - m + 1) + (2m - f^*) - (m - 1) - m = 2 - m > 0.$$

To prove the claim, observe that if $f(H_F) < \min(m, f^*)$, then x is joined to the core F_i^o by at least one edge. (Recall that if $f(H_F) < f^*$, then x is the only vertex of H_F outside F_i^o . Thus, if there is no edge between x and F_i^o , then $f(H_F) = m$.) On the other hand, b_1 is not identified with x, since this would mean that $H_B = B$, contradicting $e_0 = v_0 = v - 1$. Thus, according to the rule imposed on the attachment of B to F_i , there must be another vertex in F^* but outside F_i^o joined by an edge to F_i^o . Adding this vertex and edge to H_F results in a subgraph H' with $f(H') \le f(H_F) + m - 1$. The proof is now complete.

References

- [1] C. Berge, Graphs and Hypergraphs, North-Holland, New York (1973).
- [2] B. Bollobás, Threshold functions for small subgraphs. Math. Proc. Cambr. Phil. Soc. 90 (1981) 197-206.
- [3] P. Erdös and A. Renýi, On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17-61.
- [4] E. Győri, B. Rothschild, and A. Ruciński, Every graph is contained in a sparsest possible balanced graph. *Math. Proc. Cambr. Phil. Soc.* 98 (1985) 397-401.
- [5] M. Karoński and A. Ruciński, On the number of strictly balanced subgraphs of a random graph. Graph Theory, Lagow 1981, Lecture Notes in Mathematics 1018, Springer-Verlag, New York (1983) 79-83.
- [6] T. Łuczak and A. Ruciński, Balanced extensions of sparse graphs. Ann. Discrete Math. To appear.
- [7] C. Payan, Graphes équilibrés et arboricité rationelle. Eur. J. Combinat.
 7 (1986) 263-270.
- [8] A. Ruciński and A. Vince, Strongly balanced graphs and random graphs. J. Graph Theory 10 (1986) 251-264.
- [9] A. Ruciński and A. Vince, Balanced extensions of graphs and hypergraphs. Combinatorica 8 (1988) 279-291.
- [10] J. Spencer, Countable sparse random graphs. Random Struct. Algorithms 1 (1990) 205-214.