COMBINATORICA
Akadémiai Kiadó — Springer-Verlag

COMBINATORICA 8 (3) (1988) 279—291

BALANCED EXTENSIONS OF GRAPHS AND HYPERGRAPHS

A. RUCINSKI* and A. VINCE

Received May 28, 1985

For a hypergraph G with v vertices and e_i edges of size i, the average vertex degree is $d(G)=$ $=\Sigma ie_i/v$. Call *balanced* if $d(H)\leqq d(G)$ for all subhypergraphs H of G. Let

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

A hypergraph F is said to be a *balanced extension* of G if $G\subset F$, F is balanced and $d(F)=m(G)$, i.e. F is balanced and does not increase the maximum average degree. It is shown that for every hypergraph G there exists a balanced extension F of G. Moreover every r-uniform hypergraph has an r-uniform balanced extension. For a graph G let ext (G) denote the minimum number of vertices in any graph that is a balanced extension of G. If G has n vertices, then an upper bound of the form ext $(G)<c_1n^2$ is proved. This is best possible in the sense that ext $(G)>c_2n^2$ for an infinite family of graphs. However for sufficiently dense graphs an improved upper bound ext $(G)<c_3n$ can be obtained, confirming a conjecture of P. Erdős.

1. Introduction

A *hypergraph* G consists of a finite set $V(G)$ of vertices and a set $E(G)$ of subsets of $V(G)$ called edges. A *subhypergraph* of G, is a hypergraph whose vertex set is a subset of $V(G)$ and the edge set is a subset of $E(G)$. A hypergraph G is called *r-uniform* if each edge has size r. So a 2-uniform hypergraph is a graph. A *path* in a hypergraph is an alternating sequence $v_1e_1v_2e_2\dots v_{n-1}e_{n-1}v_n$ of vertices and edges such that each vertex belongs to the preceeding and succeeding edge. If, for each pair of vertices, there is a path joining them, then G is *connected*. A *cycle* in a hypergraph is a path with $v_1=v_n$ and $v_1, \dots, v_{n-1}$ distinct. Note that if a hypergraph is acyclic then the intersection of any two edges has cardinality at most 1. The degree of a vertex in a hypergraph is the number of edges containing v; hence the *average degree* of a hypergraph G is

$$d(G)=\frac{1}{v(G)}\Sigma\deg_{x\in V(G)} x=\frac{1}{v(G)}\Sigma ie_i(G),$$

where $v(G)$ and $e_i(G)$ denote the number of vertices and edges of size i in G. If G

* On leave from Institute of Mathematics, Adam Mickiewicz University, Poznań, Poland
AMS subject classification (1980): 05 C 35, 05 C 65.

is $r$uniform then this reduces to

$$d(G) = \frac{re(G)}{v(G)}.$$

If $d(H) \leqq d(G)$ for all subhypergraphs H of G then G is called *balanced*. Let

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

denote the maximum average degree of any subhypergraph of G. Obviously G balanced is equivalent to $m(G) = d(G)$. For graphs, this concept of balance originates in Erdős, Rényi [1] and is crucial in the investigation of random graphs [3, 5]. In Fig. 1 graph G is not balanced: $d(G) = 14/5$ and $m(G) = 3$. Graph F is balanced, contains G as a subgraph and has average degree $d(F) = 3 = m(G)$.

Fig. 1

A hypergraph F is said to be a *balanced extension* of G if $G \subset F$, F is balanced and $d(F) = m(G)$. In Fig. 1, F is a balanced extension of G. In Section 2 of this paper it is shown that every hypergraph G has a balanced extension F. Moreover for $r \geqq 2$ every r-uniform hypergraph has an r-uniform balanced extension.

Given a graph G a problem posed in [2] is to find a balanced extension F with minimum number of vertices. Let $\text{ext}(G) = \min v(F)$, the minimum taken over all graphs F that are balanced extensions of G. In Section 3 we show that $\text{ext}(G) < cn^2$, where $n = v(G)$. This upper bound is a consequence of the inductive construction used in the proof of the existence of F. Previously we thought that a smaller upper bound could be found. However, there exists a family of graphs G with $\text{ext}(G) > n^2/8$. More precisely let

$$a_n = \max_{v(G)=n} \text{ext}(G).$$

Then

$$\frac{n^2}{8} < a_n < \frac{(1+\varepsilon)n^2}{4}.$$

The examples G giving the lower bound are sparse, in the sense that $v(G) = n$, $e(G) = n+1$ and hence $d(G) = 2(1+1/n)$. In subsequent discussions, P. Erdős conjectured that for sufficiently dense graphs a tighter upper bound holds. In particular he conjectured that if the number of edges is at least cn^2, then $\text{ext}(G)$ is at most $c'n$. In Section 4 this is proved. The following question still remains open.

Problem 1.1. Is it true that if $e(G) > cn$, $c > 1$, then $\text{ext}(G) < c'n$?

2. Balanced extensions of hypergraphs

Let G be a hypergraph. Call a balanced extension F of G *uniform* if all edges in $E(F)-E(G)$ have the same size. If this common size is r, then F is called an *r-uniform balanced extension.* Note that F may be a uniform balanced extension of G, but not a uniform hypergraph

Theorem 2.1. (a) *Every hypergraph has a uniform balanced extension.*

(b) *For every $r \geqq 2$ every r-uniform hypergraph has an r-uniform balanced extension.*

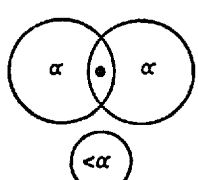
In a part (a) of the theorem it is not possible to choose, a priori, the size of the edges in the extension. More generally, let A be a finite set of natural numbers. Call a balanced extension F of G a *balanced A-extension* if for every $x \in E(F)-E(G)$, $|x| \in A$. For every A there is a hypergraph G which does not have a balanced A-extension. To see this let α be the largest element of A. If $\alpha > 1$, let G consist of two edges of size α intersecting in exactly one vertex and an isolated edge of size less than α, as in Fig. 2a. It is easy to check that G has no balanced A-extension. If $\alpha = 1$, the counterexample for $A = \{1, 2\}$ is trivially also a counterexample for $A = \{1\}$.

Let $(,)$ denotes the greatest common divisor. In the case that A has only one element the following holds.

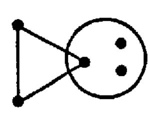
Corollary 2.2. *Let $s \geqq 2$ be an integer. Any hypergraph G with $m(G) = p/q$, $(p, q) = 1$, $(p, s) = 1$, $m(G) \geqq s/(s-1)$, has a balanced $\{s\}$-extension.*

The proof of this corollary is exactly the proof of Theorem 2.1a. In the general case we conjecture the following.

Conjecture. Let $A = \{a_1, ..., a_n\}$. If $(a_1, ..., a_n) = 1$ then every hypergraph G with $m(G) \geqq \max \{a_i/(a_i-1)\}$ has a balanced A-extension.



a)


b)

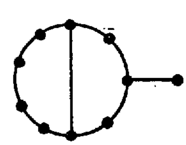

c)

Fig. 2

From the counterexample above it is clear that a lower bound on $m(G)$ is necessary. The assumptions $(p, s) = 1$ in Corollary 2.2 and $(a_1, ..., a_n) = 1$ in the conjecture are also necessary as shown by the following examples. Assume $2 | (a_1, ..., a_n)$. Consider the hypergraph G with 4 edges as shown in Fig. 2b. Then $m(G) = 2 > \max \{a_i/(a_i-1)\}$, and it can be checked that G has no balanced A-extension. Next assume $d = (a_1, ..., a_n) \geqq 3$ and let G be the graph consisting of a $(2d-1)$-cycle, a chord and an additional edge, as in Fig. 2c. Then $m(G) = 4d/(2d-1) > \max \{a_i/(a_i-1)\}$ and again it can be checked that G has no balanced A-extension.

The proof of Theorem 2.1 is inductive, the idea being that at each stage an extremely balanced hypergraph is adjoined. The existence of such a hypergraph is the subject of Theorem 2.6, which will be proved first.

Lemma 2.3. *If G is a connected r-uniform hypergraph with v vertices and e edges then* $(r-1)e \geqq v-1$.

Proof. Proceeding by induction on the number of edges, let H be a connected subhypergraph induced by $e-1$ edges of G. Then $(r-1)(e-1) \geqq v(H)-1$. Since G is connected, adding the last edge yields $r-1 \geqq v(G)-v(H)$. Hence $(r-1)e \geqq \geqq v-1$. ∎

Call an r-uniform hypergraph G *strongly balanced* if

$$\frac{e(H)}{v(H)-1} < \frac{e(G)}{v(G)-1}$$

for all non-trivial $(v(G)>1)$ subhypergraphs H of G. Note that if G is strongly balanced, then G is balanced, but not necessarily the converse. Let

$$m = \max_{H \subseteqq G} \frac{e(H)}{v(H)}$$

be called the *degree* of G and

$$m^* = \max_{H \subseteqq G} \frac{e(H)}{v(H)-1}$$

the *strong degree. Note that m differs from $m(G)$ by a factor of r.* For many applications it is convenient to work with a deficit function rather than the degree. For a hypergraph H consider a function $f(H)$ which is a linear function in $v(H)$ and $e_i(H)$, $i=1, 2, \dots$. It is easily checked that f is *modular* in the sense that for any two hypergraphs H, H'

$$f(H \cup H') = f(H) + f(H') - f(H \cap H').$$

By union and intersection of H and H' we mean the hypergraphs whose vertex and edge sets are the union and intersection, resp. of the vertex and edge sets of H and H'. The following examples of such linear functions play an important role in this paper and are called *deficit functions.* For any real number a let

$$f_a(H) = av(H) - \Sigma i e_i(H)$$

$$g_a(H) = av(H) - e(H)$$

$$h_a(H) = a\big(v(H)-1\big) - e(H).$$

The next two lemmas are direct consequences of the modularity of these deficit functions.

Lemma 2.4. (a) *A hypergraph G is balanced with $m(G)=a$ if and only if $f_a(G)=0$ and $f_a(H) \geqq 0$ for all connected subhypergraphs H of G.*

(b) *An r-uniform hypergraph G is balanced with $m=a$ if and only if $g_a(G)=0$ and $g_a(H) \geqq 0$ for all connected subhypergraphs H of G.*

(c) *An r-uniform hypergraph G is strongly balanced with* $m^*=a$ *if and only if* $h_a(G)=0$ *and* $h_a(H)\geqq 0$ *for all subhypergraphs H of G that cannot be expressed as* $H_1 > H_2$ *where* $H_1\cap H_2$ *is a single vertex.* ▮

Lemma 2.5. (a) *If G is a balanced hypergraph then each connected component* G_i *of G is balanced and* $d(G_i)=d(G)$.

(b) *If G is a strongly balanced uniform hypergraph then G is connected. Moreover, if* $G=G_1\cup G_2$ *and* $G_1\cap G_2$ *consists of a single vertex, then* G_1 *and* G_2 *are strongly balanced and have the same strong degree as G.* ▮

Given a hypergraph G we will often make use of the subhypergraph

$$\bar{G} = \bigcup_{\substack{H\subseteq G\\ f(H)=0}} H$$

where $f(H)=m(G)v(H)-\Sigma ie_i(H)$. By the modularity of f, $f(\bar{G})=0$. In other words $\bar{G}$ is the unique largest subhypergraph of G with maximum average degree.

Theorem 2.6. *For* $r\geqq 2$ *there exists a strongly balanced r-uniform hypergraph with v vertices and e edges if and only if*

$$0<\frac{v-1}{r-1}\leqq e\leqq\binom{v}{r}.$$

Proof. The necessity of the inequality follows from Lemmas 2.3 and 2.5b. The proof in the other direction is by induction on v. The result is trivially true for the hypergraph with $v=2$. Assume the theorem is true for all uniform hypergraphs with $n=v-1$ vertices. The graph case $r=2$ is proved in [4, Theorem 1]; so $r\geqq 3$ may be assumed. Four cases are considered separately.

Case 1. $e=n/(r-1)$. Let G be the unique path with e edges of size r and n vertices.

Case 2. $n/(r-1)<e\leqq n/(r-2)$. Partition n vertices into e sets $E_1', E_2', \dots, E_e'$ of sizes $r-1$ or $r-2$, the sets of size $r-2$ being exactly those E_i' with $i\in R=\{j\colon \lfloor j(x/e)\rfloor > \lfloor (j-1)x/e\rfloor,\ j=1,\dots,e\}$ where $x=e(r-1)-n$. Note that $x\geqq 1$ and $e-x\geqq 0$ and that there are x sets of size $r-2$ and $e-x$ of size $r-1$.

Adjoin a new vertex v_0; let u_i be any vertex in E_i' and let

$$E_i=\begin{cases} E_i'\cup\{u_{i+1}\} & \text{if } |E_i'|=r-1\\ E_i'\cup\{u_{i+1}\}\cup\{v_0\} & \text{if } |E_i'|=r-2,\end{cases}$$

where addition in the index is modulo e. The edges E_i form an r-uniform hypergraph G with $v=n+1$ vertices. By Lemma 2.4c, to prove that G is strongly balanced it is sufficient to check that $h_a(H)\geqq 0$ with $a=e/n$ and where H consists of $e'<e$ edges with consecutive indices $(\text{mod}\ e)$ starting with say $N+1$. Note that for any such e' edges, less than $e'x/e+1$ of them have indices in R. This is because the number of such integers in R is exactly the number of integers j such that $N(x/e) < j\leqq (N+e')x/e$. There are less than $e'x/e+1=(N+e')x/e-N(x/e)+1$ such j. Now

$$h_a(H)\geqq\frac{e}{n}\left(e'r-(e'-1)-\left(\frac{e'x}{e}+1\right)\right)-e'=\frac{e'}{n}(er-e-x-n)=0.$$

Case 3. C

$$\frac{n-1}{r-2} \leqq e \leqq \binom{n}{r-1}.$$

By the induction hypothesis let G_0 be an $(r-1)$-uniform hypergraph with n vertices and e edges. Let G be the r-uniform hypergraph formed by adjoining one vertex v_0 to G_0 and letting the edge set be $\{E\cup\{v_0\}|E\in E(G_0)\}$. It is easy to verify that G is a strongly balanced r-uniform hypergraph with v vertices and e edges.

Case 4.

$$\binom{n}{r-1} \leqq e \leqq \binom{n+1}{r}.$$

By induction let G_1 be a strongly balanced $(r-1)$-uniform hypergraph with n vertices and e_1 edges, where

$$e_1 = \max\left\{\frac{n-1}{r-2},\quad \frac{e}{n},\quad e-\binom{n}{r}\right\}.$$

For this to be possible it must be verified that $\frac{n-1}{r-1} \leqq e_1 \leqq \binom{n}{r-1}$, which is routine. Similarly let G_2 be an r-uniform hypergraph with the same n vertices as G_1 and with e_2 edges where $e_2=e-e_1$. This isposssi ble because $\frac{n-1}{r-1} \leqq e_2 \leqq \binom{n}{r}$. Now let G be the r-uniform hypergraph formed by adding one vertex v_0 to the vertex set of G_1 and G_2 and letting the edge set be

$$\{E\cup\{v_0\}|E\in E(G_1)\}\cup E(G_2).$$

It is again routine to verify that G is a strongly balanced r-uniform hypergraph with v vertices and e edges. ∎

Lemma 2.7. (a) *For any non-balanced hypergraph,* $m(G)>1$. (b) *For any r-uniform hypergraph with a cycle,* $m(G)\geqq r/(r-1)$.

Proof. (a) Let G' be a subhypergraph of G with $m(G)=d(G')$ and at least two edges. By Lemma 2.4a, G' may be assumed connected. Then $v(G')\leqq \Sigma\, ie_i(G')-1$.

(b) Let C be a cycle in G. Then

$$m(G) \geqq d(C) \geqq \frac{re(C)}{re(C)-e(C)} = \frac{r}{r-1}. \quad ∎$$

Proof of Theorem 2.1. Consider separately the case of an r-uniform hypergraph G containing no cycle. If G is connected, then G is already balanced. If not, add edges of size r to appropriate components so that $re(G_i)/v(G_i)=m(G)$ on each component G_i of G. Then the resulting r-uniform hypergraph is balanced.

Throughout the proof below r-uniform hypergraphs are assumed to have a cycle and parts (a) and (b) are proved simultaneously. Let G be a (r-uniform) hypergraph. Let $m(G)=p'/q'$ ($m=p/q$ in the uniform case), $(p', q')=1$, $(p, q)=1$. In the uniform case let $s=r$; otherwise choose s so large that $m(G)\geqq s/(s-1)$ and

$(p', s)=1$. This is possible by Lemma 2.7a. We will construct a sequence of uniform (r-uniform) extension $G \subset G_1 \subset \ldots \subset G_t$ such that $m(G)=m(G_1)=\ldots=m(G_t)$ and

$$(1) \qquad v(G_i)-v(\bar{G}_i) < v(G_{i-1})-v(\bar{G}_{i-1}), \quad i=1, \ldots, t, \quad G_0=G,$$

where t is the least index for which $G_t=\bar{G}_t$, i.e. G_t is balanced. Below we describe only the construction of G_1. The next steps are similar. Let $f^*= =\min \{f(H)\colon H \subseteqq G, f(H)>0\}$ and $f(G^*)=f^*$, where $f(H)=m(G)v(H)-\sum_i ie_i(H)$. Note that $m(G) \geqq f^*>0$. The first inequality comes from considering a subhypergraph H obtained from $\bar{G}$ by adding an isolated vertex. By the modularity of f we may assume, without loss of generality, that $G^* \supset \bar{G}$. Consider the equation

$$(2) \qquad m(G)v-se = m(G)-f^*$$

in integer variables v, e. If G is r-uniform, then dividing equation (2) by r and clearing fractions we obtain

$$(3) \qquad qe-p(v-1) = l$$

where l is an integer. Otherwise equation (2) takes the form

$$(4) \qquad q'se-p'(v-1) = l'$$

where l' is an integer. Both equations (3) and (4) have integer solutions (v, e) satisfying

$$(5) \qquad \frac{v-1}{s-1} \leqq e \leqq \binom{v}{s}.$$

Concerning $e \leqq \binom{v}{s}$, equation (2) implies $e \leqq m(G)v/s$, and $m(G)v/s \leqq \binom{v}{s}$ for sufficiently large v. Note that v can be chosen arbitrarily large because if (v, e) is a solution of (3) or (4), then so is $(v+kq, e+kp)$ or $(v+ksq', e+kp')$, resp. Concerning the inequality $e \geqq (v-1)/(s-1)$ equation (2) and the choice of s imply $e \geqq m(G)(v-1)/s \geqq (v-1)/(s-1)$. If G is r-uniform, then the above inequality holds with $s=r$ using Lemma 2.7b. Now assume that v, e are chosen to satisfy (2) and (5). By Theorem 2.6 there is an s-uniform strongly balanced hypergraph B with v vertices and e edges. Let G_1 be the hypergraph obtained by adjoining B to G so that $V(B) \cap \cap V(G)=\{x\}$, where $x \in V(G^*)-V(\bar{G})$. Equation (2) is now equivalent to

$$f(B \cup G^*) = f(B)+f(G^*)-f(\{x\}) = 0.$$

Thus $\bar{G}_1 \supseteqq B > G^*$ and so (1) holds. It only remains to show that for every $H_0 \subset G_1$ with $B_0=H_0 \cap B \neq \emptyset, f(H_0) \geqq 0$. Let $v_0=v(B_0)$ and $e_0=e(B_0)$. Because B is strongly balanced, $e_0/(v_0-1)<e/(v-1)$, which in turn, implies that $s(e-e_0)/(v-v_0)> >se/(v-1)>m(G)$; the last inequality follows from (2). Thus $f(B_0)>f(B) \geqq 0$; the last inequality follows again from (2). If $G_0=H_0 \cap G \nsubseteqq \bar{G}$ then $f(G_0) \geqq f(G^*)$. Therefore in this case

$$f(H_0) = f(B_0)+f(G_0)-f(B_0 \cap G_0) > f(B)+f(G^*)-m(G) = 0$$

by (2). If $G_0 \subseteqq \bar{G}$, then $f(H_0)=f(B_0)+f(\bar{G}) \geqq f(B) \geqq 0$. ∎

3. Balanced extensions of graphs

The subject of this and the next section is the minimum number of vertices, ext (G), in a graph that is a balanced extension of a given graph G. Let n denote the number of vertices of G. In Theorem 3.2 we prove $\text{ext}(G) \leqq (1+\varepsilon)n^2/4$. At first this seemed an extremely high estimate and a bound of the form cn was conjectured. However the graph in Fig. 3 with n vertices — consisting of an n_1-cycle, $n_1 = [n/2]$, a chord and n_2 pendant edges, $n_2 = [n/2]$, is a counterexample. If this graph is denoted by G_n, then $\text{ext}(G_n) > n^2/8$ is also proved in Theorem 3.2.

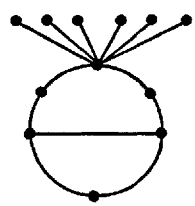


Fig. 3

Lemma 3.1. *Let G be a connected graph and S a subset of $V(G)$ with $|S|>1$ such that the distance between any two vertices of S is at least d. Then $v(G) \geqq d|S|/2$.*
Proof. For vertex $x \in S$ let $N(x) = \{y\colon d(x,y) < d/2\}$. Then the $N(x)$ are disjoint and $|B(x)| \geqq d/2$. Therefore $v(G) \geqq \sum_{x \in S} |N(x)| \geqq d|S|/2$. ▮

Recall that

$$a_n = \max_{v(G)=n} \text{ext}(G).$$

Theorem 3.2. *For any $\varepsilon > 0$ and n sufficiently large*

$$\frac{n^2}{8} < a_n < \left(\frac{1+\varepsilon}{4}\right) n^2.$$

Proof. Concerning the lower bound, let G_n be the graph on Fig. 3, Θ the subgraph consisting of the n_1-cycle and chord and V the set of pendant vertices of G_n. Let F be any balanced extension of G_n, E the set of edges in $F-\Theta$ that have a vertex in Θ, and $C_1, C_2, \ldots$ the connected components of $F-E$. Further let $V_i = V \cap V(C_i)$. Consider the case when $|V_i|=1$. Then C_i either contains a cycle or has a vertex in common with an edge in E. Otherwise F has a pendant edge y and $e(F-y)/v(F-y) > e(F)/v(F)$ contradicting the balance of F. Now F must contain a subgroup H of the form in Fig. 4a or b. In either case

$$\frac{n_1+1}{n_1} = \frac{1}{2} m(G_n) \geqq \frac{n_1+v+2}{n_1+v}$$

where n_1+v is the number of vertices in H. This implies $v(C_i) \geqq v \geqq n_1$. Next consider the $|V_i| \geqq 2$. Then for any pair x, y of vertices of V_i, F has a subgraph of the form in Fig. 4c. As above $(n_1+1)/n_1 > (n_1+v+2)/(n_1+v)$ and hence $d(x,y) \geqq$

$\geqq v-1\geqq n_1-1$. Therefore by Lemma 3.1 $v(C_i)\geqq |V_i|(n_1-1)/2$. Summing we have

$$v(F)\geqq \frac{n_2(n_1-1)}{2}+n_1\geqq \frac{n^2+2n+8}{8}.$$

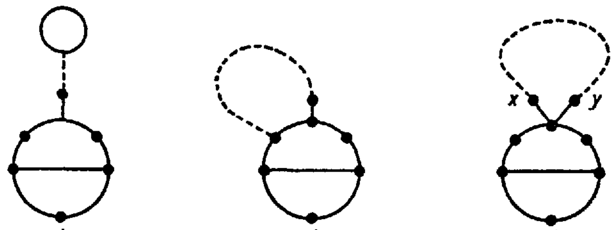


Fig. 4

Concerning the upper bound, choose any $\varepsilon>0$ and consider two cases: $m(G)\geqq \varepsilon n$ and $m(G)<\varepsilon n$. In the first case let $\bar{G}$ be as in Section 2. If $\bar{v}=v(\bar{G})$ and $\bar{e}=e(\bar{G})$, then $\bar{e}\leqq\binom{\bar{v}}{2}$ implies $\bar{v}>\varepsilon n$ and hence $e(G)\geqq\bar{e}>\bar{v}m(G)/2>\varepsilon^2n^2/2$. Now Theorem 4.2 of Section 4 applies yielding $\operatorname{ext}(G)<cn$. (For clarity of exposition, Theorem 4.2 is stated in the next section.) In the other case, $m(G)<\varepsilon n$, consider the equation

$$m(v-1)-e=-g^* \tag{1}$$

where $g^*=\min\{g_m(H)\colon H\subset G,\ g_m(H)>0\}$ and m and g_m are defined in Section 2, i.e. $m=m(G)/2$. Note that, as in the proof of Theorem 2.1, $m\geqq g^*>0$. Letting $m=p/q$ and $g^*=t/q$, $(p,q)=1$

$$qe-pv=t-p. \tag{2}$$

Note that $q\leqq\bar{v}$. Using the construction of the balanced extension F of G in Theorem 2.1

$$v(F)\leqq(n-\bar{v})v^*+\bar{v} \tag{3}$$

where (v^*,e^*) is the smallest solution of equation (1) satisfying $v^*-1\leqq e^*\leqq\binom{v^*}{2}$. If $m\geqq 1$ any solution (v,e) of equation (1) satisfies $v-1\leqq e$. If $m<1$, then G is a tree and is already balanced. Note also that for every solution of (1) with $v\geqq 2m+1$ we have $e\leqq mv\leqq\binom{v}{2}$; where the first inequality is a consequence of (1). If (v,e) is any solution of (2) then so is $(v+sq,e+sp)$. Hence

$$v^*\leqq 2m+q<\varepsilon n+q. \tag{4}$$

From (3) we have $v(F)\leqq(n-q)(\varepsilon n+q)+n$, which is maximum when $q=n(1-\varepsilon)/2$ and $v(F)\leqq n^2(1+\varepsilon)^2/4$. ∎

4. Balanced extensions of dense graphs

In the previous section it was shown that

$$\frac{n^2}{8} < a_n < \left(\frac{1+\varepsilon}{4}\right) n^2.$$

Despite this result it seems that dense graphs have balanced extensions with less than order n^2 vertices. In private communication, P. Erdős conjectured that if G has at least cn^2 edges, then G has a balanced extension with $c'n$ vertices. This is the content of the following Theorem 4.2. The graphs G in Section 3, for which $\operatorname{ext}(G) > n^2/8$, have degree $d(G) = 2(1+1/n)$. Theorem 4.2 does not apply in this case because the hypothesis of this theorem is $d(G) > cn$. The following question, stated in Problem 1.1 in the introduction, remains open: For sufficiently large graphs G with $d(G) \geqq c > 2$, is it always true that $\operatorname{ext}(G) \leqq c'n$?

The following lemma is used in the proof of Theorem 4.2. From Section 2 recall the deficit function $g(H) = mv(H) - e(H)$ where $m = e(G)/v(G)$.

Lemma 4.1. *If G is a graph such that joining two vertices of G by an edge results in a subgraph H with $g(H) \leqq 0$, then $g(G) \leqq v(G) - v(\bar{G})$.*

Proof. Assume G is such a graph. Let $H_0 = \bar{G}$ and let H_i minimize g over all subgraphs of G properly containing H_{i-1}, $i = 1, 2, \ldots$. Then $H_j = G$ for some $j \leqq v(G) - v(\bar{G})$. We prove by induction on i that $g(H_i) \leqq i$. Trivially $g(H_0) = 0$. Assume $g(H_i) \leqq i$. Adding an edge u to G with at least one end in $V(H_{i+1}) - V(H_i)$, which is always possible, must result in a subgraph $N+u$ of $G+u$ with $g(N+u) \leqq 0$. This means that $g(N) \leqq 1$ and $g(H_{i+1}) \leqq g(H_i \cup N) \leqq g(H_i) + g(N) \leqq i+1$. ∎

Theorem 4.2. *If a graph G has n vertices and more than cn^2 edges, $0 < c < 1/2$, then there is a balanced extension of G with less than $c'n$ vertices where the constant c' depends only on c.*

Proof. The idea of the proof below is to first construct a sequence of graphs $G = F_0 \subset F_1 \subset F_2 \subset \ldots \subset F_t$ such that

(i) $m(F_0) = m(F_1) = \ldots = m(F_t)$

(ii) $v(F_i) - v(F_{i-1}) < 2n \quad i = 1, 2, \ldots, t-1$

$$v(F_t) - v(F_{t-1}) < \frac{5}{2} n$$

(iii) $v(F_t) - v(\bar{F}_t) < c_1$, c_1 constant.

(iv) $t \leqq c_2$, c_2 constant.

Next the construction of Theorem 2.1b (applied to graphs) is used with at most c_1 steps. We claim

(v) Eeach such step increases the total number of vertices by less than $2n$.

Now the resulting balanced extension F of G is such that

$$v(F) \leqq v(G)+\sum_{i=1}^{t}[v(F_i)-v(F_{i-1})]+2c_1 n \leqq n+2c_2 n+\frac{1}{2}n+2c_1 n < c'n$$

for a constant c'.

The construction of F_1 is as follows (the construction of the other F is the same). Recall $m=e(G)/v(G)=p/q$, $(p,q)=1$ and let $d=\lfloor m \rfloor$. Suppose there is a set $S \subseteqq V(G)-V(\bar{G})$ with $|S|=d$. Without loss of generality it may be assumed from Lemma 4.1 that

$$g(G) \leqq v(G)-v(\bar{G}). \tag{1}$$

Adjoin to G a set W consisting of d new vertices, and join each vertex of W to all vertices of S. Call the graph obtained G_0. To G_0 add a new vertex u_1 and join it to all vertices of W. If

$$\varepsilon_1 = \min_{H \subseteqq G_0} g(H+u_1) \geqq 1,$$

(where $H+u_1$ denotes the induced subgraph) the join u_1 to any point in $\bar{G}$, thereby reducing ε_1 by1. Denote by G_1 the graph induced by G_0 and u_1. Note that $m(G_1)= =m(G_0)=m(G)$. If still $\varepsilon_1 \geqq 1$ add a new vertex u_2 and join it to all vertices of W and to any vertex in $\bar{G}$. If

$$\varepsilon_2 = \min_{H \subseteqq G_0} g(H+u_1+u_2) \geqq 1$$

add an edge from u_2 to u_1, thereby reducing ε_2 by 1. Let G_2 denote the graph induced by G_1 and u_2. Continue adding new vertices $u_3, u_4, \dots$ joining each to all vertices in W, to a vertex in $\bar{G}$ and possibly joining u_i to u_{i-1}, until for some j we have $\varepsilon_j < 1$. For the resulting graph G_j we have $m(G_j)=m(G)$ because, for any subgraph $H \subseteqq G_j$, $g(H) \geqq g(H \cap G_0) \geqq 0$ if $|V(H) \cap W| \leqq d-2$ and

$$g(H) \geqq g(H \cup \bar{G}) \geqq g((H \cup \bar{G})+u_1+\dots+u_s) \geqq 0$$

if $|V(H) \cap W| \geqq d-1$ and u_s is the last u_i in H. Using equation (1) we obtain

$$0 \leqq g(G_0+u_1+\dots+u_j) \leqq g(G_0)+(m-d)j-(2j-2) =$$

$$= g(G)+d(m-d)+(m-d)j-(2j-2) \leqq$$

$$\leqq v(G)-v(\bar{G})+d+(m-d)j-(2j-2)$$

which implies

$$j \leqq \frac{v(G)-v(\bar{G})+d+2}{2-(m-d)} \leqq v(G)-\bar{v}(G)+d+2. \tag{2}$$

Let H denote a subgraph that realizes the minimum, i.e. $g(H)=\varepsilon_j$ and let $H=H^*+ +W^*+u_1+\dots+u_j$ where $H^*=H \cap G$ and $W^*=V(H) \cap W$. Note that by the modularity of g we may assume that $\bar{G} \subseteqq H$. Now

$$|W^*| \geqq d-1; \tag{3}$$

otherwise $1 \leqq g(H^*+W^*)+1 \leqq g(H^*+W^*+u_1) \leqq g(H) < 1$. Also if $s=|V(H)\cap S|$ then $1 > g(H) \geqq g(H^*)+|W^*|(m-s)+j(m-|W^*|)-(2j-1)$ which, with equations (2) and (3) implies

$$(4) \qquad d-s \leqq \frac{v(G)-v(\bar{G})}{d}+\frac{2}{d}+1.$$

Note that $\binom{v(\bar{G})}{2}/v(\bar{G}) > e(\bar{G})/v(\bar{G}) = m$, i.e.

$$(5) \qquad v(\bar{G}) \geqq 2m+1.$$

Together with inequality (4) this implies $d-s < (n+1)/d$. Also by the assumption of the theorem

$$(6) \qquad d > m-1 \geqq \frac{e(G)}{v(G)}-1 > cn-1;$$

so if $n > \frac{1}{c}$ then

$$(7) \qquad d-s < \frac{n+1}{cn-1} < c_1$$

where c_1 is constant. Note that (6) implies $s \geqq 1$ for $n > c_1/c$. Now continue the construction of F_1. Referring to inequality (3) let W_0 be any $(d-1)$-element subset of W if $|W^*|=d$ and take $W_0=W^*$ if $|W^*|=d-1$. Add a vertex u_{j+1} to G_j and join it to all vertices of W_0 and to u_j. If $f(H+u_{j+1}) \geqq 1$ add an edge from u_{j+1} to any vertex in $\bar{G}$. Let $G_{j+1}=G_j+u_{j+1}$. Continue adding vertices $u_{j+2}, u_{j+3}, \dots$, joining each u_i to each vertex of W_0 and to u_{i-1} and possibly to a vertex in $\bar{G}$, forming G_i. Repeat until, for some l, $g(H+u_{j+1}+\dots+u_l)=0$. If $\varepsilon_j=a/q$ and $m-d=b/q$ then the number of steps $l-j$ in the above procedure is the least positive solution to the congruence $a+xb \equiv 0 \pmod{q}$. Since $(b, q)=1$ we have

$$(8) \qquad l-j < q.$$

Take $F_1=G_l$. This completes the construction. We now substantiate the claims (i)—(v) made at the beginning of the proof. First, $m(F_1)=m(G)$ is equivalent to $g(H') \geqq 0$ for all subgraphs H' of F_1. For $H' \subseteqq G_0$ this is obvious. By induction assume that for any $H' \subseteqq G_i$, $H' \nsubseteq G_0$ we have $g(H') \geqq g(H+u_{j+1}+\dots+u_i) \geqq 0$. Then for any $H'' \subseteqq G_{i+1}$ we have $g(H'')=g(H'+u_{j+1}+\dots+u_{i+1}) \geqq 0$, where $H' \subseteqq G_i$, and the claim is proved. Second, by the construction

$$(9) \qquad v(F_1)-v(G) = d+j+(l-j) \leqq d+v(G)-v(\bar{G})+d+1+q-1,$$

the inequality following from (2) and (8). Using (5) and $q \leqq n$ we have $v(F_1)-v(F_0) \leqq 2n$. Third, by the construction and (6) and (7)

$$(10) \qquad \begin{aligned} v(F_1)-v(\bar{F}_1) &\leqq v(G)-v(\bar{G})-s \\ &< v(G)-v(\bar{G})-(d-c_1) \\ &\leqq v(G)-v(\bar{G})-cn+c_1. \end{aligned}$$

In particular for $n > c_1/c$, $v(G)-v(\bar{G})$ is decreased by at least $d-c_1 > 0$. This allows the estimates (2), (4) and (9) to hold for the subsequent constructions of

$F_2, F_3, \ldots, F_{t-1}$ where t is defined as follows. Take t to be the least integer such that

$$v(F_{t-1})-v(\bar{F}_{t-1}) < d.$$

By (10) this can be accomplished with $t-1<\lceil 1/c\rceil=c_2$. To construct F_t add $d-(v(F_{t-1})-v(\bar{F}_{t-1}))$ new vertices to F_{t-1}, and call the resulting graph F'. Notice that F' can be extended to a graph F'', satisfying Lemma 4.1 (by adding edges only) and such that $\bar{F}''=\bar{F}'$. Now proceed exactly as in the construction of F_1 above to obtain F_t. By the second line of (10) $v(F_t)-v(\bar{F}_t)<c_1$. Claim (v) follows from the left side of (4) in the proof of Theorem 3.2 and (5) above. ∎

For a graph G call a balanced extension F *induced* if G is an induced subgraph of F. The proof of Theorem 1 in [4] confirms that every graph G has an induced balanced extension F with $v(F)\leqq cn^2$ where $n=v(G)$.

Problem 4.3. Does Theorem 4.2 remain valid for induced extensions?

References

[1] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), 17—61.

[2] E. Győri, B. Rothschild and A. Ruciński, Every graph is contained in a sparest possible balanced graph, *submitted.*

[3] M. Karoński, *Balanced Subgraphs of Large Random Graphs*, A. M. Univ. Press, Poznań, 1984.

[4] A. Ruciński and A. Vince, Strongly balanced graphs and random graphs, *submitted.*

[5] A. Ruciński and A. Vince, Balanced graphs and the problems of subgraphs of random graphs, *submitted.*

Andrzej Rucinski

Department of Statistics
University of Florida
Gainesville, FL 32611, U.S.A.

Andrew Vince

Department of Mathematics
University of Florida
Gainesville, FL 32611, U.S.A.