# REGULAR PARTITIONS OF GRAPHS

by

E. Szemerédi

STAN-CS-75-489
APRIL 1975

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

# REGULAR PARTITIONS OF GRAPHS

by

Endre Szemerédi

Computer Science Department
Stanford University

and

Mathematical Institute of the
Hungarian Academy of Sciences,
Budapest

Abstract

A crucial lemma in recent work of the author (showing that k-term arithmetic progression-free sets of integers must have density zero) stated (approximately) that any large bipartite graph can be decomposed into relatively few "nearly regular" bipartite subgraphs. In this note we generalize this result to arbitrary graphs, at the same time strengthening and simplifying the original bipartite result.

This research was supported in part by National Science Foundation grant GJ 36473X and by the Office of Naval Research contract NR 044-402.

We show that, for n sufficiently large, every graph with n vertices can be partitioned into k classes (k independent of n) in such a way that the resulting partition exhibits strong regularity properties. An earlier version of this result for bipartite graphs was extremely useful in proving that every set of positive integers of positive upper density contains arithmetic progressions of every length [1]. Similarly, the present version finds applications in other extremal problems of combinatorial nature. To state the result in more precise terms, we need a few definitions.

When $G = (V,E)$ is a graph and when $A$, $B$ are disjoint subsets' of $V$, we denote by $e(A,B) = e_G(A,B)$ the number of edges of G with one endpoint in A and the other in B. In addition, when A and B are nonempty, we also define

$$d(A,B) = \frac{e(A,B)}{|A|\cdot|B|} .$$

(The number $d(A,B)$ is the <u>density</u> of edges between A and B.)

The pair $(A,B)$ is called <u>$\varepsilon$-regular</u> if

$$X \subseteq A \ , \quad Y \subset B \ , \quad |X| \geq \varepsilon|A| \ , \quad |Y| \geq \varepsilon|B|$$

imply

$$|d(X,Y) - d(A,B)| < \varepsilon \qquad ;$$

otherwise the pair is called <u>$\varepsilon$-irregular</u>. By an <u>equitable partition</u> of a set $V$, we shall mean a partition of $V$ into pairwise disjoint classes $C_0, C_1, \ldots, C_k$ such that all the $C_i$'s with $1 < i < k$ have the same cardinality. The class $C_0$ may be empty; we shall refer to it as the <u>exceptional class</u>. Let $G = (V,E)$ be a graph with n vertices. An equitable partition of $V$ into classes $C_0, C_1, \ldots, C_k$ will be called s-regular if the cardinality of the exceptional class $C_0$ does not exceed $\varepsilon n$ and if at most $\varepsilon k^2$ of the pairs $(C_s, C_t)$ with $1 < s < t \leq k$ are s-irregular.

Trivially, every partition of V into me-point classes is s-regular for every $\varepsilon$ . We shall prove that for every $\varepsilon$ there is an integer M such that every sufficiently large graph admits an s-regular partition into k classes with k <M . In fact, we may also prescribe a lower bound m on the number of classes; then, of course, M becomes a function of $\varepsilon$ and m .

In the proof, we shall use the "defect form" of Schwarz inequality: if

$$\sum_{k=1}^{m} x_k = \frac{m}{n} \sum_{k=1}^{n} x_k + \delta \qquad (m \leq n)$$

then

$$\sum_{k=1}^{n} x_k^2 \geq \frac{1}{n} \left( \sum_{k=1}^{n} x_k \right)^2 + \frac{\delta^2 n}{m(n-m)} \quad .$$

We shall also use the fact that the density $d(X,Y)$ behaves in a rather continuous fashion. More precisely, if

$$X^* \subset X \ , \ Y^* \subseteq Y \ , \ |X^*| \geq (1-\delta)|X| \ , \ |Y^*| \geq (1-\delta)|Y|$$

and $0 \leq \delta < 1/3$ then some trivial computations show that

$$|d(X^*,Y^*) - d(X,Y)| < 6\delta$$

and that

$$|d(X^*,Y^*)^2 - d(X,Y)^2| < 126 \quad .$$

With every equitable partition P of the vertex-set of G into classes $C_0, C_1, \ldots, C_k$ ($C_0$ being the exceptional class), we shall associate a number called the index of P and defined by

$$\text{ind } P = \frac{1}{k^2} \sum_{s=1}^{k} \sum_{t=s+1}^{k} d(C_s, C_t)^2 \quad .$$

The crucial part of our argument is contained in the following statement.

Lemma. Let $G = (V,E)$ be a graph with $n$ vertices. Let $P$ be an equitable partition of $V$ into classes $C_0, C_1, \ldots, C_k$, the exceptional class being $C_0$. Let $\varepsilon$ be a positive integer such that

$$4^k > 600\varepsilon^{-5} .$$

If more than $\varepsilon k^2$ pairs $(C_s, C_t)$ in $1 \leq s < t \leq k$ are s-irregular then there is an equitable partition $Q$ of $V$ into $1 + k4^k$ classes, the cardinality of the exceptional class being at most

$$|C_0| + \frac{n}{4^k}$$

and such that

$$\operatorname{ind} Q > \operatorname{ind} P + \frac{\varepsilon^5}{20} .$$

Proof. For each s-irregular pair $(C_s, C_t)$ in $1 \leq s < t \leq k$, choose sets $X = X(s,t)$ and $Y = Y(s,t)$ such that

$$X \subseteq C_s \quad , \quad Y \subseteq C_t \quad , \quad |X| \geq \varepsilon |C_s| \; , \; |Y| \geq |C \quad |$$

and

$$|d(X,Y) - d(C_s, C_t)| > \varepsilon \quad .$$

In each $C_i$, those sets define the obvious equivalence relation with at most $2^{k-1}$ classes; the equivalence classes will be called the atoms. Set

$$m = \left\lfloor \frac{1}{4^k} |C_i| \right\rfloor \qquad (1 \leq i \leq k) \quad .$$

Trivially, we may choose a collection $Q$ of pairwise disjoint subsets of $V$ such that

(i) every member of Q has cardinality m ,

(ii) every atom A contains exactly $\lfloor |A|/m \rfloor$ members of Q ,

(iii) every class $C_i$ contains exactly $\lfloor |C_i|/m \rfloor$ members of Q .

Note that

$$\lfloor |C_i|/m \rfloor = 4^k \qquad (1 \le i \le k)$$

and so every class $C_i$ contains exactly $4^k$ members of Q ; in fact, we may assume that Q has exactly $k4^k$ members. The collection Q may be considered to be an equitable partition of V , the cardinality its exceptional class being at most

$$|C_0| + km \le |C_0| + \frac{n}{4^k} \quad .$$

It remains to be shown that

$$\text{ind } Q > \text{ind } P + \frac{\varepsilon^5}{20}$$

For this purpose, label all the members of Q which are contained in some $C_s$ $(1 \le s \le k)$ as $C_s(i)$ with i running from 1 to $q = 4^k$ . For each s , define

$$C_s^* = \bigcup_{i=1}^{q} C_s(i) \quad .$$

Then

$$|C_s^*| > |C_s| - m > |C_s| \left( 1 - \frac{\varepsilon^5}{600} \right)$$

and so

$$\left| d(C_s^*, C_t^*)^2 - d(C_s, C_t)^2 \right| < \frac{\varepsilon^5}{50}$$

whenever $1 \le s < t \le k$ . Now, the Schwarz lemma implies

$$\frac{1}{q^2}\sum_{i=1}^{q}\sum_{j=1}^{q} d(C_s(i)\ ,\ C_t(j))^2 > d(C_s^*, C_t^*)^2 > d(C_s,C_t)^2 - \frac{\varepsilon^5}{50} .$$

The last inequality can be greatly improved whenever the pair $(C_s,C_t)$ happens to be s-irregular. In this case, we shall make use of the sets $X = X(s,t)$ and $Y = Y(s,t)$ introduced above. Let $X_0$ be the largest subset of $X$ that partitions into members of $Q$ . Evidently,

$$|X_0| \geq |X| - 2^k m > |X|(1 - \frac{\varepsilon}{100}) .$$

We shall set

$$r = \left\lceil \frac{|X|}{m}\left(1 - \frac{\varepsilon}{100}\right)\right\rceil .$$

Without loss of generality, we may assume that

$$X^* = \bigcup_{i=1}^{r} C_s(i) \subseteq X \quad \text{and} \quad Y^* = \bigcup_{j=1}^{r} C_t(j) \subseteq Y .$$

We have

$$|X^*| \geq |X|\left(1 - \frac{\varepsilon}{100}\right) , |Y^*| \geq |Y|\left(1 - \frac{\varepsilon}{100}\right)$$

and so

$$|d(X^*,Y^*) - d(X,Y)| < \frac{\varepsilon}{4} .$$

Hence

$$|d(X^*,Y^*) - d(C_s^*, C_t^*)| > \frac{\varepsilon}{2} .$$

Using the defect form of Schwarz inequality (with $n = q^2$ , $m = r^2$ and $8 = r^2 d(X^*,Y^*) - r^2 d(C_s^*, C_t^*))$ we obtain

$$\frac{1}{q^2}\sum_{i=1}^{q}\sum_{j=1}^{q} d(C_s(i)\ ,C_t(j))^2 \geq d(C_s^*, C_t^*)^2 + \frac{\varepsilon^2}{4} \cdot \frac{r^2}{q^2-r^2}$$

$$> d(C_s,C_t)^2 - \frac{\varepsilon^5}{50} + \frac{\varepsilon^4}{16} .$$

Finally, we have

$$\text{ind } Q \geq \frac{1}{k^2} \sum_{s=1}^{k} \sum_{t=s+1}^{k} \left( \frac{1}{q^2} \sum_{i=1}^{q} \sum_{j=1}^{q} d(C_s(i), C_t(j))^2 \right)$$

$$\geq \frac{1}{k^2} \sum_{s=1}^{k} \sum_{t=s+1}^{k} \left( d(C_s, C_t)^2 - \frac{\varepsilon^2}{50} \right) + \frac{1}{k^2} \varepsilon k^2 \cdot \frac{\varepsilon^4}{16}$$

$$> \text{ind } P + \frac{\varepsilon^5}{20}$$

as desired.

<u>Theorem</u>. For every positive real $\varepsilon$ and for every positive integer $m$ there are positive integers $N$ and $M$ with the following property: for every graph $G$ with at least $N$ vertices there is an e-regular partition of $G$ into $k+1$ classes such that $m \leq k \leq M$ .

<u>Proof</u>. Let $s$ be the smallest integer such that

$$4^s > 600\varepsilon^{-5}, \quad s > m \quad \text{and} \quad s \geq \frac{2}{\varepsilon}$$

Define a sequence $f(0), f(1), f(2), \ldots$ by setting $f(0) = s$ and

$$f(t+1) = f(t) 4^{f(t)}$$

for every $t$ . Let $t$ be the largest nonnegative integer for which there exists an equitable partition $P$ of $V$ into $1 + f(t)$ classes such that

$$\text{ind } P \geq \frac{t\varepsilon^5}{20}$$

and the size of the exceptional class does not exceed

$$\varepsilon n \left( 1 - \frac{1}{2^{t+1}} \right)$$

(Such a partition certainly exists for $t = 0$. Since ind $P \le \frac{1}{2}$ for every partition $P$, the integer $t$ is well defined.) By our lemma, and by the maximality of $t$, the partition $P$ is $\varepsilon$-regular. Hence we may set $M = f(\lfloor 10\varepsilon^{-5} \rfloor)$. End of proof.

It would be interesting to decide whether the same statement holds when the requirement that at most $\varepsilon k^2$ pairs $(C_s, C_t)$ be $\varepsilon$-irregular is replaced by the stronger requirement that no pairs $(C_s, C_t)$ be $\varepsilon$-irregular.

Acknowledgment

My indebtedness to my friend Vaclav Chvatal is extremely great. In fact, he wrote this paper after listening to my rough oral presentation.

References

[1] Endre Szemeredi, "On a set containing no k elements in arithmetic progression,", Acta Arithmetica, to appear.