References.

1. H. Hadwiger, "Eine Erweiterung eines Theorems von Steinhaus-Rademacher", *Commentarii Mathematici Helvetici*, 19 (1946–47), 236–239.
2. H. Kestelman, "Convergent sequences belonging to a set", *Journal London Math. Soc.*, 22 (1947), 130–135.

Peterhouse,
Cambridge.

A THEOREM ON PLANAR GRAPHS

PETER UNGAR*.

Definitions. A graph G consists of a set of elements $N_1, \ldots, N_f$ which we call nodes and a set of different pairs of nodes $(N_i N_j)$, $i \neq j$, called edges. For our purpose it is convenient not to exclude graphs in which the set of edges is empty.

A subgraph of G is a graph whose nodes and edges all belong to G.

Let H be a subgraph of G. Then $G-H$ denotes the graph consisting of those nodes of G which do not belong to H and those edges of G which have neither end in H.

$f(G)$ denotes the number of nodes in G.

G is called planar if its nodes N can be represented by points in a plane, and the edges $N_i N_j$ by curves joining N_i to N_j, in such a way that these curves do not intersect.

The main purpose of this paper is to prove the following Theorem 1. I found this theorem while considering the question "How many non-adjacent countries can be selected from a map of f countries in the plane?".

THEOREM 1. *Let G_0 denote a planar graph. Let k be a positive integer > 1. There exists a set of nodes I such that the components of $G_0 - I$ contain at most k nodes each and*

$$f(I) < 12k^{-\frac{1}{2}}(\log k)^{\frac{3}{2}} f(G_0).$$

As far as I can see the proof could not be simplified by aiming at $f(I) = o(1) f(G_0)$ instead of the explicit expression in Theorem 1. The theorem would probably remain true if we replaced the factor $12 (\log k)^{\frac{3}{2}}$ by a constant. It can be shown by considering the "hexagonal lattice" that this constant would be $\geqslant \sqrt{3}$.

Proof of Theorem 1. For $2 \leqslant k \leqslant 10^5$ Theorem 1 is trivial, because in this interval $12k^{-\frac{1}{2}} \log^{\frac{3}{2}} k > 1$. We suppose $k > 10^5$.

* Received 23 May, 1950; read 15 June, 1950.

LEMMA 1. *Let $a_1, \ldots, a_\delta$ be positive integers with sum f. If*

$$\delta > 2\log^2 f,\ f > 10^5, \tag{1}$$

then there exists a ν such that

$$\min\left(\sum_1^{\nu-1} a_\mu,\ \sum_{\nu+1}^{\delta} a_\mu\right) > \frac{\delta}{2\log f} a_\nu.$$

Proof. We may suppose without loss of generality that $\sum_1^{[\frac{1}{2}\delta]} a_\mu \leqslant \frac{1}{2}f$. Then for $\nu \leqslant \frac{1}{2}\delta$ the first sum in the min is the smaller. Suppose now that Lemma 1 is false. The function $A(x) = a_{[x+1]}$ then satisfies the following inequalities:

$$A(x) \geqslant 1 \text{ and } A(x) \geqslant 2\delta^{-1}\log f\left(\int_0^x A(u)\,du - A(x)\right) \text{ in } 0 \leqslant x \leqslant [\tfrac{1}{2}\delta];$$

or if we put
$$L = 2\log f/(\delta + 2\log f),$$

$$A(x) \geqslant L\int_0^x A(u)\,du.$$

The function

$$\alpha(x) = 1 \text{ in } 0 \leqslant x \leqslant L^{-1} \text{ and } \alpha(x) = e^{Lx-1} \text{ for } L^{-1} \leqslant x \leqslant [\tfrac{1}{2}\delta]$$

satisfies

$$\alpha(x) \leqslant A(x) \text{ for } 0 \leqslant x \leqslant L^{-1} \text{ and } \alpha(x) = L\int_0^x \alpha(u)\,du \text{ for } L^{-1} \leqslant x \leqslant [\tfrac{1}{2}\delta].$$

Hence $\alpha(x) \leqslant A(x)$ in $0 \leqslant x \leqslant [\frac{1}{2}\delta]$. But since $L < \frac{1}{12}$, by (1),

$$\int_0^{[\frac{1}{2}\delta]} \alpha(x)\,dx = L^{-1}e^{L[\frac{1}{2}\delta]-1} \geqslant 12e^{\frac{1}{2}L[\delta-1]-1} > 3e^{\frac{1}{2}L\delta}$$
$$= 3\exp\left[\log f/(1+2\delta^{-1}\log f)\right] > 3\exp\left(\log f - 2\delta^{-1}\log^2 f\right) > f;$$

whereas
$$\int_0^{[\frac{1}{2}\delta]} A(x)\,dx < \sum_1^{\delta} a_\mu = f,$$

which is a contradiction.

Definitions. The *distance* of two nodes in a graph is the number of edges that a shortest way connecting them contains.

The *diameter* $d = d(G)$ of a connected graph G is the smallest number with the property that the distance between any two nodes of G is $\leqslant d$.

LEMMA 2. *Let the connected graph G be such that $d(G) \geqslant 2\log^2 f(G)$, $f(G) > 10^5$. Then there exists a non-empty set of nodes I such that no component of $G - I$ contains more than*

$$f(G) - (d+1)f(I)/\{2\log f(G)\}$$

nodes.

Proof. Let P and Q be two nodes at a distance d from each other. Let a_μ be the number of nodes at distance $\mu-1$ from P. These a_μ satisfy the requirements of Lemma 1 with $\delta=d+1$. If we choose ν according to Lemma 1, then the nodes lying at a distance $\nu-1$ from P form a set I with $f(I)=a_\nu$. The number g of nodes of any component of $G-I$ satisfies

$$g\leqslant \max\left(\sum_1^{\nu-1} a_\mu, \sum_{\nu+1}^{d+1} a_\mu\right) < f(G)-\min\left(\sum_1^{\nu-1} a_\mu, \sum_{\nu+1}^{d+1} a_\mu\right)$$
$$< f(G)-(d+1)f(I)/\{2\log f(G)\}.$$

LEMMA 3. *Let G be a connected planar graph with diameter d. There exists a set of nodes I, such that $f(I)\leqslant 2d+1$ and no component of $G-I$ contains more than $\frac{2}{3}f(G)$ nodes.*

Proof. For $d\geqslant \frac{1}{6}f(G)$ the result is trivial. We therefore suppose $d<\frac{1}{6}f(G)$.

Let us call a planar graph saturated if no edge can be added to it without destroying the planar property of the graph. It is enough to prove Lemma 3 for saturated G. For, if we have any graph we can add edges to it until we get a saturated graph G. A set I which satisfies Lemma 3 for G, will also satisfy Lemma 3 for the original graph, because the addition of edges does not increase d.

Let G be a saturated graph. We imagine it to be drawn on the plane. The domains D into which G decomposes the plane are triangles.

Definition. A *geodetic line* is a way between two nodes containing as few edges as possible. There may of course be several g-lines between two nodes.

We now define a class of subgraphs of G which we call Y-graphs. Let A and B in G be connected by an edge and let O be a third node. Let $\overline{OA}$ and $\overline{OB}$ be arbitrarily chosen but fixed geodetics from O to A and B respectively, such that, if F denotes the node farthest from O belonging to both $\overline{OA}$ and $\overline{OB}$, part $\overline{OF}$ of $\overline{OA}$ is identical with part $\overline{OF}$ of $\overline{OB}$. If $A\neq F\neq B$, we say that the edges and nodes of $\overline{OF}$, together with the edges and nodes inside or on the boundary of the Jordan curve consisting of $\overline{FA}$, edge AB and $\overline{FB}$, form a Y-graph. If $F=A$ or $F=B$, the corresponding Y-graph is defined as the edges and nodes of the geodetics $\overline{OA}$ and $\overline{OB}$, one of which forms part of the other. If F differs from O, A and B, the boundary of the graph resembles the shape of the letter Y.

The following argument will show the existence of a Y-graph with $\frac{1}{3}f(G)<f(Y)\leqslant\frac{2}{3}f(G)$.

We start with a Y-graph Y_1 with $f(Y_1)>\frac{2}{3}f(G)$. Such a Y_1 exists; G itself is a Y-graph with the three nodes on the outer boundary taking the roles of O, A, B. Y_1 is not a single geodetic, for that would give

$d+1 \geqslant f(Y_1) > \frac{2}{3}f(G)$. So the Jordan curve $\overline{FA}$, AB, $\overline{BF}$ exists. One of the two triangles which have the edge AB on their boundary, contains interior points of the Jordan curve. The third node C of this triangle belongs to Y_1.

Take any geodetic from O to C. Let F' be the last node it has in common with either $\overline{OA}$ or $\overline{OB}$ (possibly $F' = C$). Replace the part OF' of the geodetic by the part OF' of $\overline{OA}$ or $\overline{OB}$. We get a new geodetic from O to C; denote it by $\overline{OC}$. Every edge and node of $\overline{OC}$ belongs to Y_1. The geodetics $\overline{OA}$, $\overline{OC}$ and $\overline{OB}$, $\overline{OC}$ define two new Y-graphs Y_2 and Y_2'. Every node of Y_1 belongs to at least one of them. The edge AB however belongs to neither.

As $f(Y_1) > \frac{2}{3}f(G)$, either $f(Y_2') > \frac{1}{3}f(G)$ or $f(Y_2) > \frac{1}{3}f(G)$. Suppose the latter. If $f(Y_2) \leqslant \frac{2}{3}f(G)$, we have the Y-graph we want. If $f(Y_2) > \frac{2}{3}f(G)$, we can repeat the construction. We get a sequence of Y-graphs each containing fewer edges than the previous one. Hence the sequence terminates. Its last member Y_i will satisfy $\frac{1}{3}f(G) < f(Y_i) \leqslant \frac{2}{3}f(G)$.

The set I of nodes belonging to the geodetics defining the graph Y_i separate the other nodes of Y_i from the nodes not belonging to Y_i. A largest component of $G-I$ clearly contains less than $\frac{2}{3}f(G)$ nodes. Also, by definition, the two geodetics to which the nodes of I belong contain not more than $d+1$ nodes each and O belongs to both. Hence $f(I) \leqslant 2d+1$.

This proves Lemma 3.

We construct now a sequence of disjoint sets of nodes $I_0, I_1, \ldots, I_q$, such that $I = I_0 + \ldots + I_q$ satisfies Theorem 1.

Suppose I_p has been defined for all $-1 \leqslant p < h$. I_{-1} is the empty set. G_0 is the whole graph. For $h > 0$, let G_h denote a largest component of $G_0 - I_0 - \ldots - I_{h-1}$, *i.e.* an arbitrarily chosen but fixed component containing not less nodes than any other component of $G_0 - I_0 - \ldots - I_{h-1}$. If $f(G_h) \leqslant k$, put $q = h-1$. Let $f(G_h) > k$. We put $\lambda(G_h) = \{\frac{1}{3}f(G_h)\log f(G_h)\}^{\frac{1}{2}} - 1$. If $d(G_h) < \lambda(G_h)$, we apply Lemma 3 to G_h and obtain I_h. If $d(G_h) \geqslant \lambda(G_h)$, we apply Lemma 2 to G_h and obtain I_h. Lemma 2 can be applied because $f(G_h) > k > 10^5$ and $d(G_h) \geqslant \lambda(G_h) > 2\log^2 f(G_h)$. Since $f(I_h) \geqslant 1$ for $h \geqslant 0$ and the sets I_h are disjoint, the construction ends in a finite number of steps.

We have to estimate $f(I)$. Let G_h' denote a largest component of $G_h - I_h$. If $d(G_h) < \lambda(G_h)$, we have, by Lemma 3, $f(I_h) < 2\big(d(G_h)+1\big)$ and

$$f(G_h) > f(G_h') + \tfrac{1}{3}f(G_h)$$

$$= f(G_h') + \frac{\lambda(G_h)+1}{\sqrt{3}}\sqrt{\left(\frac{f(G_h)}{\log f(G_h)}\right)} > f(G_h') + \frac{1}{2\sqrt{3}}f(I_h)\sqrt{\frac{f(G_h)}{\log f(G_h)}};$$

or

$$f(I_h) \leqslant 2\sqrt{3}\,.\,f(G_h - G_h')\sqrt{\left(\frac{\log f(G_h)}{f(G_h)}\right)}. \tag{2}$$

If $d(G_h) \geqslant \lambda(G_h)$, we get from Lemma 2

$$f(G_h) \geqslant f(G_h')+\frac{d(G_h)+1}{2\log f(G_h)} f(I_h) \geqslant f(G_h')+\frac{1}{2\sqrt{3}} \sqrt{\left(\frac{f(G_h)}{\log f(G_h)}\right)} \cdot f(I_h).$$

This again implies (2), so that (2) is true for all h.

Now let x be any positive number. Denote by S the set of indices h for which

$$x \leqslant f(G_h - G_h') < 2x. \tag{3}$$

Then
$$\sum_{h \in S} f(G_h - G_h') \leqslant 2f(G_0). \tag{4}$$

Proof of (4). (4) will be proved if we show that if l, m, n are in S, $l < m < n$, then no node can belong to all three graphs $G_l - G_l'$, $G_m - G_m'$, $G_n - G_n'$. Suppose the node N belongs to all three graphs. $N \notin I_l$; otherwise N could not belong to G_m, because $m > l$. Denote by G_l'' the component of $G_l - I_l$ which contains N. As $G_l'' \in G_l - G_l'$, we have

$$f(G_l'') \leqslant f(G_l - G_l') < 2x.$$

Now G_l'' is the component of $G_0 - I_0 - \ldots - I_l$ containing N. G_m is a component of $G_0 - I_0 - \ldots - I_{m-1}$ and also contains N. As $m > l$, G_m is a subgraph of G_l'' and $f(G_m) \leqslant f(G_l) < 2x$. As above, $N \notin I_m$. Denote by G_m'' the component of $G_m - I_m$ which contains N. G_m' and G_m'' are disjoint and both subgraphs of G_m. By definition of G_m', $f(G_m'') \leqslant f(G_m')$. So $f(G_m'') \leqslant \frac{1}{2} f(G_m) < x$. Again, G_n is a subgraph of G_m'', and therefore

$$f(G_n - G_n') < f(G_n) \leqslant f(G_m'') < x.$$

This means $n \notin S$, contrary to our hypothesis. Thus (4) is proved.

We divide the indices $0 \leqslant h \leqslant q$ into classes S_m. $h \in S_m$ if (3) holds with $x = 2^{m-2} 3^{-\frac{1}{2}} k^{\frac{1}{2}} (\log k)^{-\frac{1}{2}}$. Every h belongs to a class with $m \geqslant 1$. For it follows from (2), together with $f(I_h) \geqslant 1$ and $f(G_h) > k$ $(0 \leqslant h \leqslant q)$, that

$$f(G_h - G_h') \geqslant \frac{1}{2\sqrt{3}} \sqrt{\left(\frac{f(G_h)}{\log f(G_h)}\right)} \geqslant \frac{1}{2\sqrt{3}} \sqrt{\left(\frac{k}{\log k}\right)},$$

because $\sqrt{(u/\log u)}$ is an increasing function for $u > e$.

From (2) and (4),

$$\sum_{1 \leqslant m \leqslant \log k} \sum_{h \in S_m} f(I_h) < \sum_{1 \leqslant m \leqslant \log k} \sum_{h \in S_m} 2\sqrt{3}\, f(G_h - G_h') \sqrt{\left(\frac{\log f(G_h)}{f(G_h)}\right)}$$

$$< \sum_{1 \leqslant m \leqslant \log k} \sum_{h \in S_m} 2\sqrt{3} \, . f(G_h - G_h') k^{-\frac{1}{2}} (\log k)^{\frac{1}{2}}$$

$$< \sum_{1 \leqslant m \leqslant \log k} 4\sqrt{3} f(G_0) k^{-\frac{1}{2}} (\log k)^{\frac{1}{2}} < 7 f(G_0) k^{-\frac{1}{2}} (\log k)^{\frac{3}{2}}. \tag{5}$$

For $m > \log k$ we use the fact that, if $h \varepsilon S_m$,

$$f(G_h) > f(G_h - G_h') \geqslant 2^{m-2} 3^{-\frac{1}{2}} k^{\frac{1}{2}} (\log k)^{-\frac{1}{2}} > 2^{m-[\log k]} \cdot \tfrac{1}{5} k \tag{6}$$

for $k > 10^5$.

We put $m - [\log k] = \mu$. We get, from (2), (6) and (4),

$$\sum_{m > \log k} \sum_{h \varepsilon S_m} f(I_h) < \sum_{\mu = 1, 2, \ldots} \sum_{h \varepsilon S_m} 2\sqrt{3} \cdot f(G_h - G_h') \{\log (2^\mu k)\}^{\frac{1}{2}} \left(\frac{2^\mu k}{5}\right)^{-\frac{1}{2}}$$

$$< \sum_{\mu=1}^{\infty} 4\sqrt{3} \cdot f(G_0) \sqrt{5} \{\log(2^\mu k)\}^{\frac{1}{2}} (2^\mu k)^{-\frac{1}{2}}. \tag{7}$$

Now

$$\sum_{\mu=1}^{\infty} \{\log(2^\mu k)\}^{\frac{1}{2}} (2^\mu k)^{-\frac{1}{2}} < \int_0^\infty (x \log 2 + \log k)^{\frac{1}{2}} e^{-\frac{1}{2} x \log 2 - \frac{1}{2} \log k} dx$$

$$= \int_{\log k}^{\infty} u^{\frac{1}{2}} e^{-\frac{1}{2} u} \frac{du}{\log 2}$$

$$= \frac{2}{\log 2} (\log^{\frac{1}{2}} k) k^{-\frac{1}{2}} + \frac{1}{\log 2} \int_{\log k}^{\infty} u^{-\frac{1}{2}} e^{-\frac{1}{2} u} du$$

$$< 3k^{-\frac{1}{2}} (\log^{\frac{1}{2}} k + \log^{-\frac{1}{2}} k).$$

Substituting this into (7) and adding (5), we obtain

$$f(I) = \sum_{m = 1, 2, \ldots} \sum_{h \varepsilon S_m} f(I_h)$$

$$< 7k^{-\frac{1}{2}} (\log k)^{\frac{3}{2}} f(G_0) + 12\sqrt{15} \cdot k^{-\frac{1}{2}} (\log^{\frac{1}{2}} k + \log^{-\frac{1}{2}} k) f(G_0)$$

$$< 12 k^{-\frac{1}{2}} (\log k)^{\frac{3}{2}} f(G_0)$$

if $k > 10^5$, which establishes the theorem.

The planar property of G was used only in the proof of Lemma 3. The question arose whether a similar result could be proved for other classes of graphs. I considered regular graphs of degree 3 (*i.e.* graphs in which 3 edges concur at every node) and found the following negative result.

THEOREM 2. *Let k be a positive integer. There exist regular graphs R of degree 3 with the following property: if I is any set of nodes such that the components of $R - I$ each contain less than k nodes, then*

$$f(I) > \tfrac{1}{4} f(R).$$

We prove Theorem 2 by showing:

(*a*) There exist regular connected graphs R of degree 3 with arbitrarily many nodes which contain no circuit of length $< k$. In Tutte's terminology* the girth of R is $\geqslant k$.

(*b*) Any graph of degree 3 and girth $\geqslant k$ satisfies Theorem 2.

* *Proc. Camb. Phil. Soc.*, 43 (1947), 459–474. (*a*) was also proved some years previously by Mr. W. T. Tutte, but not published.

Proof of (*a*). *Definition.* The *distance* of two edges e, e' is one less than the number of edges in a shortest geodetic with e, e' as ends.

For $k=1$, (*a*) is trivial. Suppose (*a*) is proved for $k=n-1$. We want to prove it for $k=n$ and $f(R)>c>2^n$, where c is an arbitrarily large number. By the inductive hypothesis there exists a connected regular graph T of degree 3 with $f(T)>c$ and girth $n-1$. Let e be an edge in T which belongs to at least one circuit of length $n-1$. Let e' be an edge at distance $n-2$ from e. Such an edge exists, because the number of edges at distance i from e is at most 2^{i+1} and the number of edges at distance $<n-2$ from e is therefore at most $\sum_{i=0}^{n-3} 2^{i+1} < 2^n < f(T)$, whereas the number of all the edges in T is $\frac{3}{2}f(T)$. Now put two new nodes E, E' on e, e' respectively and connect them by an edge. We get a new graph of degree 3; call it S. Let the five new edges that S contains in place of e, e' be e_1, e_2, EE', e_1', e_2'. None of these belongs to a circuit of length $n-1$, for the following reasons.

If EE' belonged to a circuit of length $n-1$ in S, then there would be in T a geodetic with $n-2$ edges and e, e' as ends, contrary to hypothesis. So any $n-1$ circuit containing e_1 also contains e_2. Such a circuit would correspond to an $n-2$ circuit in T. This is also contrary to hypothesis. Similarly for e_1', e_2'.

S contains at least one less edge belonging to an $n-1$ circuit than T. $f(S)>f(T)>c$. We can repeat the process until all edges belonging to $n-1$ circuits are eliminated, and we get a graph as required in (*a*).

Proof of (*b*). Let I be as described in Theorem 2. Let R_i $(i=1, 2, \ldots)$ be the components of $R-I$. As $f(R_i)<k$, R_i contain no circuits. By an obvious property of trees, the number of edges in R_i is $f(R_i)-1$. Hence there are $3f(R_i)-2\big(f(R_i)-1\big)=f(R_i)+2$ edges in R_i, with one end-node in I and the other in R_i. Summing over i, we find that there are $\sum_i \big(f(R_i)+2\big)$ edges in R with one end belonging to I. The nodes being of degree 3 in R, this gives

$$3f(I) \geq \sum_i \big(f(R_i)+2\big) > \sum_i f(R_i) = f(R-I);$$

or

$$4f(I) > f(R).$$

To Dr. R. Rado I wish to express my thanks for the many improvements he suggested. Mr. G. A. Dirac obliged me by pointing out a gap in an earlier version of the proof of Lemma 3.

University College,
London.