GRÖTZSCH'S THEOREM ON 3-COLORINGS

Branko Grünbaum

1. INTRODUCTION

Grötzsch [3] established the following remarkable result: *Every planar graph with no* 3-*circuits is* 3-*colorable.* The aim of the present paper is to establish the following generalization of Grötzsch's result:

THEOREM. *Every planar graph with not more than three* 3-*circuits is* 3-*colorable.*

This result is in a certain sense best possible, since there exist infinitely many planar graphs with four 3-circuits which are not 3-colorable. However, we were unable to prove the conjecture: *If a planar graph* G *is not* 3-*colorable, then* G *contains two pairs of (edge or vertex) incident triangles.*

The present paper arose from an attempt to find a simple proof of Grötzsch's theorem. Berge's [1] endeavor to give a short proof failed since the claim, essential to his proof, that every planar graph without 3-circuits is a subgraph of a rigid-circuit graph (see Dirac [2]) containing no complete 4-graph, is invalidated by simple counterexamples (such as, for example, the net of a cube).

Our attempt did yield a proof somewhat simpler than Grötzsch's original proof, although the last step (Section 5 of the present paper) is only a reformulation of his ingenious arguments. (There are other similarities between the proofs, but they involve rather standard arguments.) The simplified proof allows one, however, to prove Grötzsch's theorem in the generalized form given above.

In the present paper, only graphs without 1- or 2-circuits are considered. A graph is said to be k-colorable if its nodes may be assigned to k classes in such a way that no two nodes belonging to the same class be connected by an edge of the graph. Without loss of generality, we shall assume that all the graphs considered are 2-connected. A 2-connected planar graph may have different imbeddings in the 2-sphere (for 3-connected graphs the faces are uniquely determined); we shall, however, always work with a *given* imbedding of the graph, and therefore, with a well-defined set of faces determined on the sphere by the graph. For nodes A_m and A_n of a circuit C, we denote by $d(A_m, A_n)$ the number of edges in the shorter arc of C connecting A_m and A_n. We shall use the terms *triangle, quadrangle,* and *pentagon* for 3-face, 4-face, 5-face, respectively.

2. PROOF OF THE THEOREM

We begin by establishing

PROPOSITION 1. *Let* G *be a planar graph. Then there exists a planar graph* H *with the following properties:*

Received November 21, 1962.

The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research, OAR through the European Office, Aerospace Research, United States Air Force.

(i) G *is a subgraph of* H,

(ii) *no face of* H *has more than* 5 *edges,*

(iii) *every* 3*-circuit of* H *belongs to* G.

Proof. The graph H is obtained from the given realization of G by repeating the following procedure: Let K be a k-gonal face of G, $k \geq 6$. If there is an edge in G connecting non-neighbouring nodes A_i, A_j, of K, we choose two nodes, A_m, A_n, of K that are separated by A_i, A_j, and such that $d(A_m, A_n) \geq 3$; taking any point A* in the interior of K, we adjoin to the graph the node A* and the edges A^*A_m and A^*A_n. If, on the other hand, G contains no edge connecting non-neighbouring nodes of K, we choose any two non-neighbouring nodes A_m, A_n, of K and any point A* in the interior of K, and adjoin the node A* and the edges A^*A_m and A^*A_n. In either case, K is split into two faces, each with less than k edges, and obviously no new triangles are introduced. This ends the proof of Proposition 1.

As a consequence of Proposition 1, it is obvious that we may assume, in the proof of the Theorem, that each face of G is either a triangle, or a quadrangle, or a pentagon. For the inductive proof, we find it technically advantageous to prove the following more complicated statement:

PROPOSITION 2. *Let* G *be a planar graph having as faces only triangles, quadrangles, and pentagons and containing at most three* 3*-circuits. Then it is possible to color* G *with* 3 *colors. Moreover, if* G *contains at most one* 3*-circuit, the colors on the nodes of one arbitrarily chosen* 4*- or* 5*-face may be prescribed, unless the chosen face is a pentagon three consecutive vertices of which form a* 3*-circuit—in which case the prescribed coloring of the pentagon must assign different colors to those three vertices.*

Proposition 2 will be proved by induction on the number of edges of G, the assertion being obvious for graphs with a small number of edges. We shall describe different *reductions*, that is, constructions indicating how a given G is to be 3-colored to satisfy Proposition 2, assuming the validity of Proposition 2 for graphs with a smaller number of edges. The set of reductions is chosen in such a way as to cover all possible cases. This will complete the proof of Proposition 2 and thus also the proof of the Theorem.

The first reduction applies to graphs G that have a k-cut, $k \leq 5$, where a k-cut of G is a (simple) k-circuit which is not a k-face of G (therefore each of the connected components of the complement (in the sphere) of a k-cut of G contains at least one edge of G).

For graphs G that are free of k-cuts, $k \leq 5$, we describe reductions applicable in the following cases:

(i) G contains a node of valence 2,

(ii) G contains a quadrangle,

(iii) G contains no quadrangle and no node of valence 2, but contains a triangle having a node of valence 3,

(iv) G is not covered by (i), (ii), (iii); that is, G contains no quadrangle, no node of valence 2 and, if any triangles are present, all their nodes have valence of at least 4.

The details of the different reductions will be given in Sections 3, 4, and 5. We conclude the present section by explaining the terminology used in the sequel.

While the 3-coloring of a triangle is unique, a quadrangle can be 3-colored either by assigning the same color to diametric points, or by assigning to one pair of diametric points one color, the members of the other pair being colored by different colors. The 3-coloring of a pentagon is also unique, up to a cyclic permutation of its vertices; two pairs of vertices are colored by two colors, while the remaining vertex (*special vertex* in the sequel) receives the third color.

For graphs G with at most one 3-circuit, the quadrangle or pentagon whose coloring is prescribed will be called the *distinguished face* of G.

If C is a k-cut of G, we shall call the *part* (of G) *determined by* C either of the two graphs obtained from G by deleting the nodes and edges of G contained in one of the connected components of the complement of C in the sphere. Note that C is a face for each of the parts. A graph with no 3-, 4-, or 5-cuts will be called *cut-free*.

For a planar graph (imbedded in the sphere), we shall denote by v, e, f, and f_k, the number of nodes, edges, faces, and k-faces, respectively, of G.

The *identification* of two nodes N_1 and N_2 of a graph G consists in omitting from G the two non-neighbouring nodes N_1 and N_2 (and the edges incident to them) and introducing a new node N connected to all nodes of G that were connected to at least one of the nodes N_1 and N_2. From any coloring of the reduced graph, a coloring of G is derived by assigning to both N_1 and N_2 the color of N.

3. REDUCTIONS FOR GRAPHS WITH k-CUTS, $k \leq 5$

If G contains a 3-cut T, the reduction is immediate: each of the two parts of G determined by T has less edges than G; therefore, each of the parts is 3-colorable, and since the coloration of T is unique, the parts join along T in a proper way.

If G has no 3-cuts but has a 4-cut Q, we distinguish two cases:

(i) $f_3 = 2$ *or* $f_3 = 3$. One of the parts determined by Q contains at most one triangle. We color first the other part, and then the first one, with Q as distinguished face.

(ii) $f_3 = 0$ *or* $f_3 = 1$. The distinguished face D is contained in one of the parts determined by Q; we 3-color this part first, and use Q as distinguished face for the other part.

If G has no 3-cuts and no 4-cuts but has a 5-cut P, the procedure again depends on f_3.

(i) $f_3 = 0$ *or* $f_3 = 1$. As above, the distinguished face D is contained in one of the parts determined by P; we 3-color this part first, and 3-color the other part with P as the distinguished face. The only possible exception is if $f_3 = 1$, the distinguished face D is contained in one of the parts determined by $P = \{ABCDE\}$, while the other part contains the only triangle present in the *exceptional configuration* (see Fig. 1). Since G has no 3- or 4-cuts, T and Q are faces. If the nodes B and D, or C and E (see Fig. 1), are joined in G by a path of length ≤ 3 different from those in Fig. 1, G has a 5-cut which is not of the exceptional kind, which can be used instead of P. If, on the contrary, B and D, and C and E, are not joined by such paths, at least one of these pairs can be identified without introducing new 3-circuits and without interfering with the coloring of the distinguished face D.

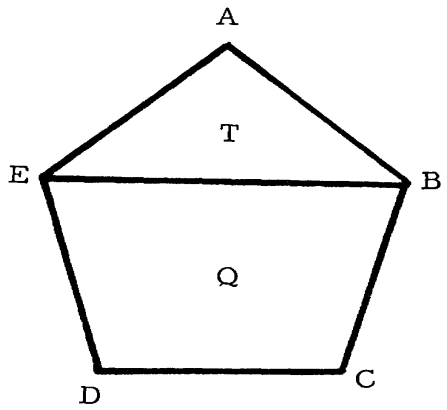
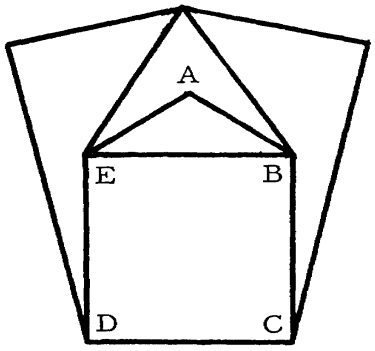


Figure 1

(ii) $f_3 = 2$. If both triangles belong to one of the parts determined by P, we 3-color this part first and use P as distinguished face for the other part. If P separates the triangles, we use P as distinguished face for both parts, taking a 3-coloring of P which is compatible with the occurrence of the exceptional configuration of Fig. 1 on either or both sides of P. (Note that there always exists such a 3-coloring of P.)

(iii) $f_3 = 3$. If all three triangles are in one of the parts determined by P, we 3-color this part first and use P as distinguished face for the other part. In the only other case, one of the parts of G determined by P contains one triangle, and the other two triangles. We 3-color first the latter part, and use P as distinguished face for the remaining part. The only possible complication (analogous to the exceptional case in (i) above) arises if the part containing the one triangle is the configuration of Fig. 1. In this case we attempt to reduce G by identifying B and D, or C and E. If the identification of B and D introduces a new triangle (and is therefore unacceptable), there exists a 3-path connecting B and D, disjoint from P. If a similar situation were to prevail regarding C and E, G would contain one of the configurations in Figs. 2 and 3. The configurations of Fig. 2 are ruled out by the assumption that G



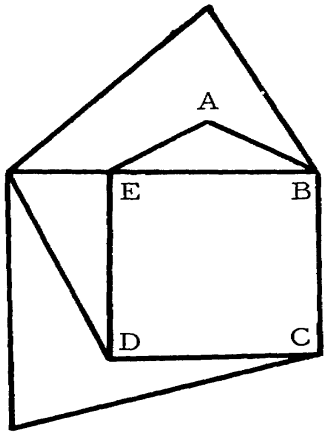


Figure 2

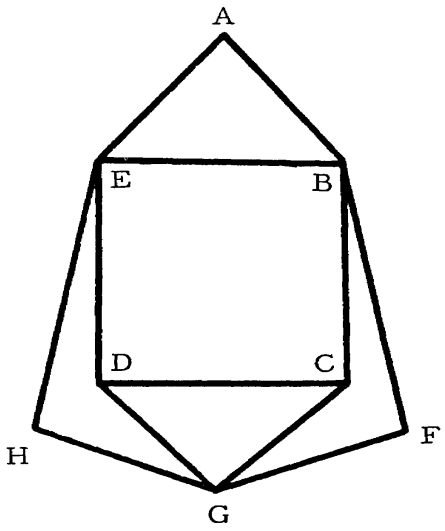


Figure 3

has no 3- or 4-cuts. As for the configuration of Fig. 3, it contains a 5-cut which is not of the exceptional kind. Indeed, both BCGHE and BFGDE are 5-cuts. If either of them were the 5-circuit of the exceptional configuration in Fig. 1, G would contain a 3- or 4-cut, contrary to the assumption.

Thus if G contains a k-cut with $k \leq 5$, one of the above reductions applies, and there remains to be considered the case of graphs having no such cuts. Note that this implies, in particular, that no triangle has an edge in common with either a triangle or a quadrangle. (The configuration of Fig. 1 is the only exception, but for it the assertion of Proposition 2 is obvious.)

4. REDUCTIONS OF CUT-FREE GRAPHS

Throughout this section, G will denote a graph without 3-, 4-, or 5-cuts. We shall describe reductions applicable in the cases (i), (ii), and (iii) of Section 2.

(i) Let G contain a node N of valence 2, and let N_1 and N_2 be the neighbours of N. We omit N and adjoin the edge $N_1 N_2$ (if it is not already contained in G). Note that, unless G is one of the graphs of Fig. 4 for which Proposition 2 obviously holds, the two faces of G incident to N are pentagons. The reduced graph being 3-colored, N is assigned the color different from those of N_1 and N_2. This procedure is inapplicable only in the following cases:

(a) G contains a vertex $N^* \neq N$ and edges $N^* N_1$ and $N^* N_2$,

(b) N belongs to a pentagonal distinguished face P and N is neither the special node of P nor a neighbour of the special node of P.

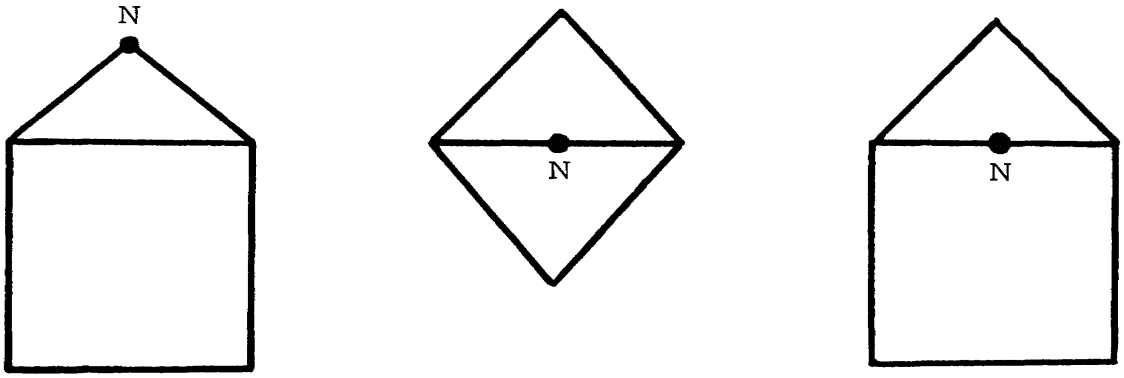


Figure 4

In case (a), G would have a 4- or 5-cut, contrary to the assumption. In case (b), in which $f_3 \leq 1$, we omit N and identify the nodes N_1 and N_2. Since G is cut-free, this introduces at most two additional triangles (but eliminates the distinguished face).

Thus, if a cut-free G has a node of valence 2, G is reducible.

(ii) Let G be cut-free and contain no node of valence 2, and let Q be a 4-face of G. If Q is not the distinguished face D of G, there exists a node N_1 of Q which does not belong to D. We form a reduced graph G* by identifying N_1 with the node N_3 opposite to N_1 in Q. (Note that no new 3-circuits arise since G is cut-free.) The 3-coloring of G* yields a 3-coloring of G, the nodes N_1 and N_3 receiving the same color.

The above procedure yields a reduction for all graphs with $f_4 > 0$, with one exception: $f_4 = 1$, $f_3 \leq 1$, and the only 4-face Q is the distinguished face. In this case there exists an edge $N_1 N_2$ of Q which is incident to a pentagon P (Fig. 5). We omit the edge $N_1 N_2$, introduce the new edge $N_0 N_2$, and take in the resulting graph Q* the pentagon $\{N_0 N_2 N_3 N_4 N_1\}$ as the distinguished face, with a 3-coloring coinciding on $\{N_1 N_2 N_3 N_4\}$ with the one prescribed for Q. The graph Q* contains at least one 4-face and is not of the exceptional type; therefore, one of the previous reductions applies.

Thus, all cut-free graphs with $f_4 > 0$ are reducible.

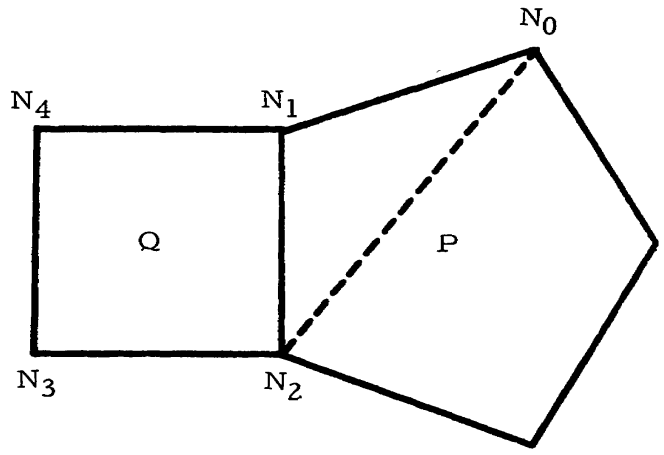


Figure 5

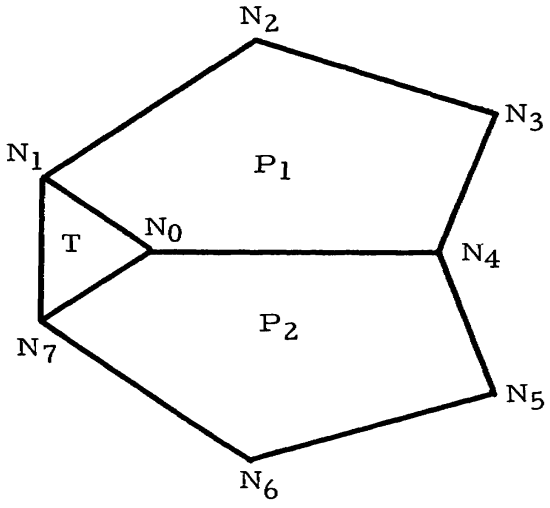


Figure 6

(iii) Let G be a cut-free graph with no node of valence 2, such that $f_4 = 0$, $f_3 > 0$, and at least one of the triangles has a node N_0 of valence 3 (Fig. 6). We take a pentagon incident to N_0 which is not the distinguished face; assume this to be P_2. Then we omit N_0 and identify N_4 and N_7; let G* be the resulting graph. The number of 3-circuits in G* is the same as that in G. Clearly, a 3-coloring of G* yields a 3-coloring of G, except if P_1 is the distinguished face and the special node is N_2 or N_3. In that case we apply a different reduction: a graph G** is formed from G by omitting N_0 and adjoining the edge N_4N_7. G** contains no 3-circuits, and we 3-color it, taking $\{N_1N_2N_3N_4N_7\}$ as the distinguished face with a 3-coloring that agrees on N_1, N_2, N_3, N_4 with that prescribed for P_1.

This completes the reduction procedures for the case (iii).

5. PROOF OF PROPOSITION 2 (END)

We have still to describe reductions for graphs G satisfying the following conditions (graphs of type (P)):

(a) G has no k-cuts, $k \leq 5$,

(b) every node of G has valence of at least 3,

(c) every face of G is either a triangle or a pentagon,

(d) every node incident to a triangle has valence of at least 4.

This case will be disposed of in three steps. First, the existence of a special configuration of pentagons in each graph of type (P) will be established; next, two possible reductions will be described for that configuration; lastly, it will be shown that at least one of the reductions is always applicable.

Given any planar graph G, we assign (following Lebesgue [4] and Grötzsch [3]) to each node N of G the weight $w(N) = 1/k$, where k is the valence of N. To any face F of G, we assign the weight $w(F) = \Sigma w(N)$, where N runs over all the nodes incident to F. Clearly, $\Sigma w(F)$, where the summation is over all the faces F of G, equals v, the number of nodes of G. If G is a graph of type (P), then, as is easily checked, $w(T) \leq 3/4$ for any triangle T in G, and $w(P) \leq 3/2$ for any pentagon P

in G that is not of the following special type: four vertices of P have valence 3, the fifth has valence 3, 4 or 5. For such special pentagons $3/2 < w(P) \leq 5/3$.

Since for graphs of type (P), $f = f_3 + f_5$, $2e = 3f_3 + 5f_5$, and Euler's formula $v + f = e + 2$ holds, we obtain (denoting by f_5^* the number of pentagons of the special type)

$$\frac{1}{2}f_3 + \frac{3}{2}f_5 + 2 = v = \sum_F w(F) \leq \frac{3}{4}f_3 + \frac{5}{3}f_5^* + \frac{3}{2}(f_5 - f_5^*);$$

that is, $24 \leq 3f_3 + 2f_5^*$. But $f_3 \leq 3$, and therefore $f_5^* \geq 15/2$; that is $f_5^* \geq 8$. Thus, each graph of type (P) contains at least 8 pentagons P with four vertices of valence 3; that is, the configuration of Fig. 7 obtains. Therefore, we may choose among them one in which neither P, nor any P_i (i = 0, 1, 2, 3, 4) is the distinguished face of G.

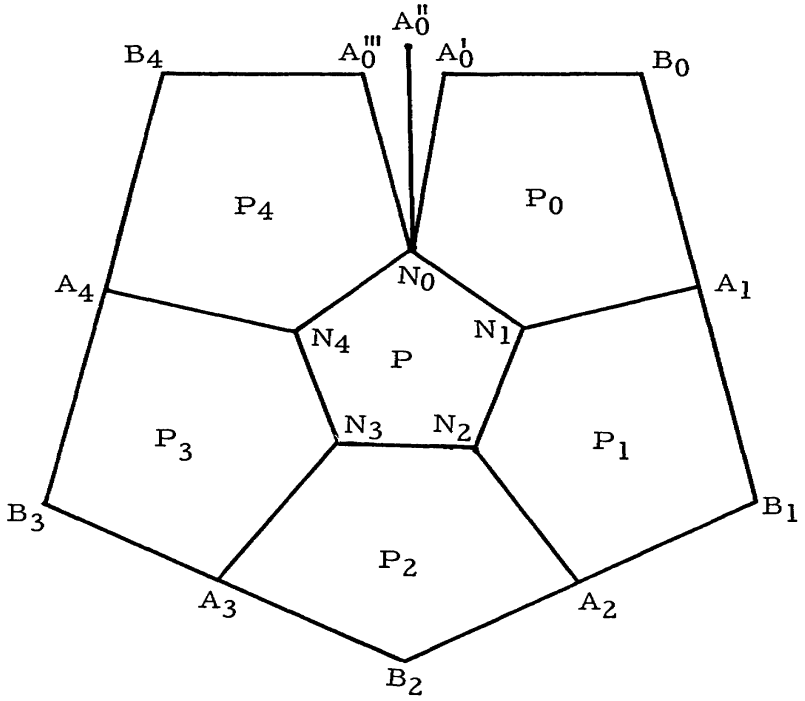


Figure 7

Note that A_0'' possibly coincides with one or both of the nodes A_0' and A_0'''; but, since G is of type (P), no other coincidences may occur among the nodes in Fig. 7.

The two reductions (due to Grötzsch [3]) of the configuration of Fig. 7 are as follows. In both, the nodes N_1, N_2, and N_3 are omitted. In the first reduction, the node N_4 is identified with A_1, and the node A_2 is identified with A_3. Thus the configuration of Fig. 8 is obtained. The second reduction consists in identifying the node N_0 with A_3 and the node A_1 with A_2, yielding the configuration of Fig. 9.

It is easily checked that if the graph obtained from G on replacing the chosen configuration of Fig. 7 by that of Fig. 8 or that of Fig. 9 is 3-colored, it is possible to 3-color G by assigning to N_1, N_2, N_3 appropriate colors. Thus if either of the two reductions is possible, the proof of Proposition 2 is complete.

The only obstacle to either of the reductions is the appearance of 3-circuits. Since $d(A_1, A_2) = d(A_2, A_3) = 2$, the identification of A_1 with A_2, or that of A_2

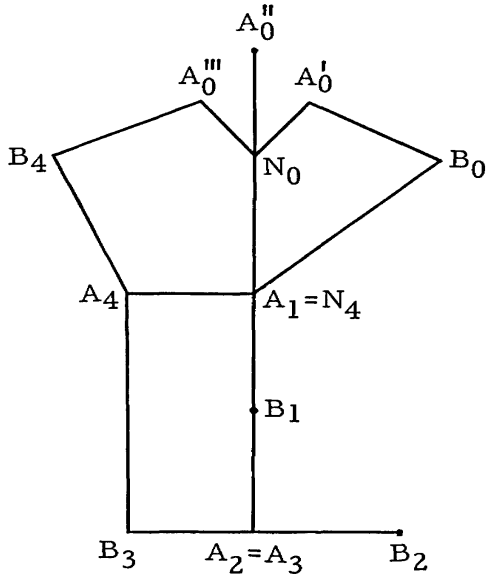


Figure 8

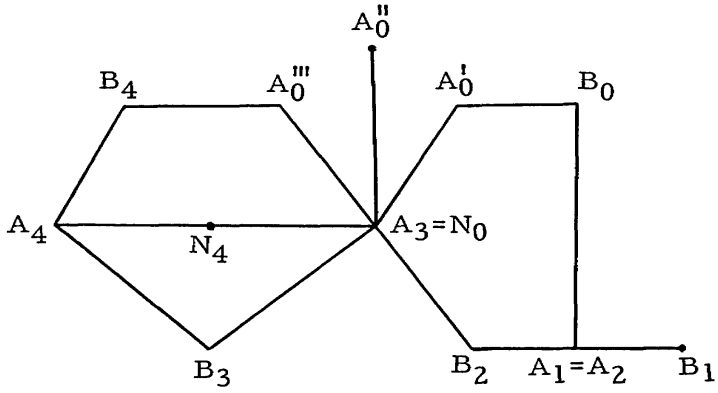


Figure 9

with A_3, does not cause the appearance of additional 3-circuits (G being of type (P)).

Thus any 3-circuit arising in the first reduction must pass through $A_1 = N_4$ and, since $d(N_0, A_1) = d(N_3, A_1) = 2$, through A_4. Therefore, G contains a node A* and edges A^*A_1, A^*A_4; A* may coincide with B_1 or with B_3, but not with B_0 or B_4 since G has no 5-cuts.

Similarly, any 3-circuit arising in the second reduction must pass through $A_3 = N_0$ and one of the nodes A_0', A_0'', A_0'''. Also, G must contain a node A** and edges $A^{**}A_3$ and one of $A^{**}A_0'$, $A^{**}A_0''$, $A^{**}A_0'''$. The node A** may coincide with one of the nodes B_0 or B_2.

If both reductions were to lead to 3-circuits, it would follow that A* = A** (and therefore, it would be different from all the nodes B_i), and a contradiction would result through the presence of a 4-circuit $A^*A_3B_4A_4$. Thus at least one of the above reductions is possible, and the proof of Proposition 2 is completed.

REFERENCES

1. C. Berge, *Les problèmes de coloration en théorie des graphes*, Publ. Inst. Statist. Univ. Paris, 9 (1960), 123-160.

2. G. A. Dirac, *On rigid circuit graphs*, Abh. Math. Sem. Univ. Hamburg, 25 (1961), 71-76.

3. H. Grötzsch, *Zur Theorie der diskreten Gebilde*, VII. *Ein Dreifarbensatz für dreikreisfreie Netze auf der Kugel*, Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg. Math.-Mat. Reihe, 8 (1958/59), 109-120.

4. H. Lebesgue, *Quelques conséquences simples de la formule d'Euler*, J. Math. Pures Appl., (9) 19 (1940), 27-43

The Hebrew University
Jerusalem