stichting
mathematisch
centrum

AFDELING INFORMATICA
(DEPARTMENT OF COMPUTER SCIENCE)

IW 81/77 JUNI

P.M.B. VITÁNYI

HOW "GOOD" CAN A GRAPH BE n-COLORED?

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): 05C15, 68A20

ACM-Computing Reviews-categories: 5.32, 5.25

How "good" can a graph be n-colored? *)

by

Paul M.B. Vitányi

ABSTRACT

The problem of how "near" we can come to an n-coloring of a given graph is investigated. I.e., what is the minimum possible number of edges joining equicolored vertices if we color the vertices of a given graph with n colors. In its generality the problem of finding such an optimal color assignment to the vertices (given the graph and the number of colors) is NP-complete. For each graph G, however, colors can be assigned to the vertices in such a way that the number of offending edges is less than or equal to the total number of edges divided by the number of colors. Furthermore, an $\mathcal{O}(epn)$ deterministic algorithm for finding such an n-color assignment is exhibited where e is the number of edges and p is the number of vertices of the graph ($e \geq p \geq n$). A priori solutions for the minimal number of offending edges are given for complete graphs; similarly for equicolored K_m in K_p and equicolored graphs in K_p.

KEY WORDS & PHRASES: *graph coloring, construction of course schedules with minimization of conflicts, computational complexity, analysis of algorithms, NP complete.*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

In order to facilitate meetings at international summit conferences, simultaneous translations ought to be provided for each pair of nonequilingual participants. What is the maximal number of translators possibly needed for p persons, each speaking one out of n distinct tongues, under the assumption that everybody may talk to everybody else at the same time. (Situations like that seem to be one of the reasons why large meetings function badly - even without the language barrier). The question asks for the maximally possible number of edges connecting distinct colored vertices in "n-colorings" of K_p, the complete graph on p vertices. In the sequel we shall assume the graph theoretical terminology and notation of HARARY [4].

An n-coloring of a graph G = (V,E) is a partition of V in disjoint subsets $V_1,V_2,\ldots,V_n$ such that no element of E belongs to $\bigcup_{i=1}^{n} V_i \times V_i$. The minimum number n such that an n-coloring is possible is called its *chromatic number* $\chi(G)$. Given some partition π of V into $V_1,V_2,\ldots,V_n$ the ratio $\#\{a \in E \mid a \in \bigcup_{i=1}^{n} V_i \times V_i\} / \# E$ is a measure of how far we are removed from an n-coloring of G. E.g., if $\chi(G) = n' \leq n$ then there is a partition π such that this ratio is 0. (Obviously, there always is a partition such that the ratio is 1, viz. $V_1 = V$ and $V_2,V_3,\ldots,V_n = \emptyset$). In this paper we investigate the following question: given a graph G and an integer n what is the infimum of the above ratio and for what partition π is it reached if the partition ranges over all possibilities. Therefore we introduce the following:

DEFINITION. Let G = (V,E) be a graph and let π be a partition of V in disjoint subsets $V_1,V_2,\ldots,V_n$. The number of *bad* edges in G with respect to π is

$$B(G,\pi) = \#\{a \in E \mid a \in \bigcup_{i=1}^{n} V_i \times V_i\}$$

and the number of *good* edges in G is

$$C(G,\pi) = \#E - B(G,\pi).$$

The *n-colorability distance* of G is

$$d(G,n) = \inf_{\pi} B(G,\pi) \; / \; \#E$$

and the *n-colorability* of G is

$$c(G,n) = 1 - d(G,n).$$

Clearly, $c(G,1) = 0$ and $c(G,\chi(G)) = 1$; $c(G,n)$ is strictly increasing on the interval $[1,\chi(G)]$ and $c(G,n) = 1$ for $n \geq \chi(G)$. It is well known that $c(G,5) = 1$ for all planar graphs G and by the claimed proof of Haken and Appel of the four color conjecture $c(G,4) = 1$ for all planar graphs G.

The notions outlined above have been used (at least implicitly) in the construction of course schedules and minimization of conflicts therein (a conflict being represented by an edge joining courses which cannot be held simultaneously). However, there usually heuristic methods are elaborated while we propose to look at the problem from a more theoretical point of view. In the following we show that finding $c(G,n)$ is NP complete and give an $\mathcal{O}(p\{{p \atop n}\})$ deterministic algorithm to find $c(G,n)$ and the corresponding vertex partition where $\{{p \atop n}\}$ is a Stirling number of the second kind. We exhibit an $\mathcal{O}(epn)$ deterministic algorithm which for a graph G with e edges and p vertices always finds an n-partition of the vertices such that the ratio of bad edges to the total number of edges is less than or equal to $1/n$. A priori solutions to $c(G,n)$ are given for the case that G is a complete graph and similar problems about the number of equicolored K_m in K_p and equicolored graphs in K_p are investigated.

2. n-COLORABILITY FOR ARBITRARY GRAPHS

To determine c in general is hard; indeed, it is NP-complete. That is to say, it is solvable by a nondeterministic algorithm in time polynomial in the number of vertices (i.e., the algorithm can "guess" the partition yielding c) and every problem which can be solved by a nondeterministic algorithm in time polynomial in the "size" of the problem can be reduced to this one deterministically in time proportional to a polynomial in the "size" of the problem. This latter fact follows since the problem of determining the chromatic number of a graph is NP-complete, and determining c(G,2), c(G,3), ,..., c(G,n), with c(G,n) = 1 and c(G,n') < c(G,n) for n' < n, yields $\chi(G) = n$. For further details about NP-complete problems of this nature see KARP [6]. The problem of determining c(G,n) can be stated as a quadratic programming problem, LENSTRA [7], and can be interpreted as a time-table problem. It can be approximated rather easily (by single vertex color changes) to local optima in the space of all n-partitions of the vertex set but to find a global optimum in the general case presumably requires checking all possibilities. Refraining from doublures like isomorphic partitions etc., it is easily seen that checking all elements in

$$W = \{1\{1\}^{i_1}2\{1,2\}^{i_2}3\ldots n\{1,2,\ldots,n\}^{i_n} \mid i_1 + i_2 + \ldots + i_n = p - n \text{ and } i_1, i_2, \ldots, i_n \geq 0\}$$

suffices, the interpretation being that each element of the above set denotes a different assignment of n colors to the ordered set of p vertices of the graph G. If we denote the cardinality of the above set by S(p,n) we see that

$$S(p,n) = \sum_{\substack{i_1 + i_2 + \ldots + i_n = p - n \\ i_1, i_2, \ldots, i_n \geq 0}} 1^{i_1} 2^{i_2} \ldots n^{i_n}.$$

By expansion of generating functions we note that S(p,n) is generated by

$$\frac{1}{(1-z)(1-2z)\ldots(1-nz)} = \sum_{p=n}^{\infty} S(p,n) z^{p-n}$$

and by [5, ch.IV]

$$\frac{z^n}{(1-z)(1-2z)\ldots(1-nz)} = \sum_{p=n}^{\infty} \left\{{p \atop n}\right\} z^p$$

where $\left\{{p \atop n}\right\}$ is a Stirling number of the second kind. Hence $S(p,n) = \left\{{p \atop n}\right\}$ and since these Stirling numbers can be expressed as:

$$\{{p \atop n}\} = \frac{(-1)^n}{n!} \sum_{i=0}^{n+1} (-1)^i \binom{n}{i} i^p$$

we infer that $\{{p \atop n}\} < {n^p}/{n!}$. (For further facts concerning Stirling numbers see the above reference.) In [1] an $O(\{{p \atop n}\})$ algorithm is described for generating the sequence of n-partitions W by one or two digit changes (changes in the color assignment of a single vertex or two vertices in the graph). Hence an algorithm to determine c(G,n) works as follows (where we write the algorithms in self-explanatory pidgin ALGOL):

ALGORITHM I.

Initiate with partition π_0 and compute $B(G,\pi_0)$;

Min B := $B(G,\pi_0)$; d_1 := d_2 := 0; π := π_0;
while not all partitions have been generated
  do generate new partition π_{i+1} from π_i
    using PARTEXACT from [1];
    $d_{1,2}$:= number of bad edges incident
      on the one or two vertices changing
      color in going from π_i to π_{i+1},
      in π_i and π_{i+1}, respectively;
    $B(G,\pi_{i+1})$:= $B(G,\pi_i) - d_1 + d_2$;
    if Min B > $B(G,\pi_{i+1})$
    then (Min B := $B(G,\pi_{i+1})$; π := π_{i+1})
    fi
  od
¢ c(G,n) := 1 − Min B/#E and π contains the corresponding n-color assignment ¢.

Since there are at most p − 1 edges incident on a vertex, computing d_1 and d_2 takes $O(p)$ time; the condition in the if statement can be satisfied at most $O(p^2)$ times since the number of bad edges is decreased by at least 1 each time. Therefore we have (since PARTEXACT runs in time $O(\{{p \atop n}\})$) a total time of $O(p)O(\{{p \atop n}\}) + O(p^2)O(p)O(n)$ for the algorithm. Hence:

THEOREM 1. *There is an algorithm which determines* c(G,n) *for a graph* G *on* p *vertices in (deterministic) time* $O(p\{{p \atop n}\})$.

At this point we should contrast our approach with that of GAREY and JOHNSON [3]. Faced with a computationally intractable problem, it is a practical approach to relax the optimality constraint. Instead of requiring an optimal coloring one might, for instance, be willing to settle for a coloring which uses "close" to the optimal number of colors. It has been shown in [3], however, that to find an n-coloring of a graph G is still NP-hard for all $n \leq r\,\chi(G) + d$ for constant d and $r < 2$. We have relaxed the optimality constraint in the other way. Faced with the problem of finding an n-coloring of G, where n might or might not be sufficient for finding a "true" n-coloring we settle for an assignment of the given n colors to the vertices of G such that the number of offending edges is minimal. We have shown that this problem too is NP-hard. It will now appear that, given G and n, we can always find a color assignment for which the number of bad edges is less than or equal to $1/n$-th of the total number of edges e in G and that there is an $\mathcal{O}(epn)$ deterministic algorithm to find this color-assignment.

<u>LEMMA</u> 2. $d(G,n) \leq 1/n$ *for all graphs* G.

<u>PROOF</u>. Let a vertex v of color j be connected by x_i edges with vertices of color i, $1 \leq i,j \leq n$. Let $x_{min} = \min_{1 \leq i \leq n} \{x_i\}$. As long as $x_j > x_{min}$ for some vertex v the number of bad edges can be decreased by changing the color of v to i for some $x_i = x_{min}$. Hence we can assume that for a partition yielding $d(G,n)$ all vertices are colored in such a way that the number of bad edges incident on each vertex v is less than or equal to degree(v)/n, which proves the Lemma. □

That Lemma 2 cannot be improved upon is seen by, e.g., the example of figure 1.

The next algorithm embodies the procedure implicit in the previous proof: it finds a color assignment π, using n colors, for a given graph G such that $B(G,\pi)/\#E(G) \leq 1/n$. The color assignment found is a local optimum, i.e., it cannot be improved by changing the color of a single vertex.

<u>ALGORITHM</u> II. (A,p,n). ¢ We assume that G is available as a vertex adjacency matrix A[1:p,1:p] of dimension $p \times p$ and that n colors are used. Furthermore, we dispense with declarations. ¢

```
initialize: for i from 1 by 1 to p
            do last := 1; π[i] := 1;
                for j from 1 by 1 to p
                do if j ≤ n then D[i,j] := 0 fi;
                   if A[i,j] ≠ 0
                   then (A[i,last] := j; last := last + 1)
                   fi
                od;
                D[i,1] := degree[i] := last - 1;
                Min[i,1] := 0; Min[i,2] := 2
            od; i := 1; check := p;
```

¢ Now A[i,j], $1 \le i \le p$ and $1 \le j \le$ degree[i] contains the number of the j-th vertex connected with vertex i. π[i] contains the color of vertex i and all vertices are colored initially with color 1. D[i,c] contains the number of edges connecting vertex i with vertices of color c; Min[i,1] := $\inf_{\substack{1\le i\le p\\ 1\le c\le n}}$ D[i,c] and Min[i,2] the associated color c. ¢

```
Loop: while i ≠ check
      do while D[i,π[i]] > Min[i,1]
         do c := π[i]; π[i] := Min[i,2];
            for j from 1 by 1 to degree[i]
            do D[A[i,j],c] := D[A[i,j],c] - 1;
               D[A[i,j],π[i]] := D[A[i,j],π[i]] + 1;
               if Min[A[i,j],2] = c
               then Min[A[i,j],1] := Min[A[i,j],1] - 1
               else if Min[A[i,j],2] = π[i]
                    then if D[A[i,j],k] = inf{D[A[i,j],s]}
                                         1≤s≤n
                         then (Min[A[i,j],1] := D[A[i,j],k]; Min[A[i,j],2] :=k)
                         fi
                    fi
               fi
            od;check := i
         od;
         i:= if i = p
            then 1 else i + 1
            fi
      od;
```

¢ The loop checks cyclically over all vertices whether a change in color of a single vertex can decrease the number of bad edges; if so all relevant variables are updated in accordance with the new color assignment. If the loop is executed p times without the main condition being satisfied the program terminates. The array π contains the resulting color assignment and the number of bad edges is $\sum_{\substack{1\le j\le n \\ \pi[i]=j}} D[i,j]/2$. ¢

Time analysis of Algorithm II: initialize runs in time $O(p^2)$. The condition in the main loop can be fulfilled at most e = #E(G) times since the number of bad edges is decreased by at least 1 each time. Each vertex has degree $O(p)$. Looking for the infimum of n elements takes at most n comparisons. Hence the total algorithm runs in $O(p^2) + O(e)O(p)O(n)$. Assuming without loss of generality that $e > p > n$ we have:

<u>THEOREM</u> 3. *There is an algorithm which colors the* p *vertices of a given graph* G *with* n *colors such that for the resulting partition* π *it holds that* $B(G,\pi)/e \le 1/n$, *with* $e = \#E(G)$, *and which runs in deterministic time* $O(epn)$.

Clearly, Algorithm II can only be guaranteed to find local optima of $B(G,\pi)/\#E(G)$ (with variable π) with respect to single vertex color changes, in view of the fact that to find d(G,n) is NP-hard.

It should be noted that for finding an optimal n color assignment to the vertices of a graph G only those vertices have to be considered which have degree ≥ n, since the vertices of lesser degree can always be colored so that they have no incident bad edges. Therefore, we assume that G has more than np/2 edges, and Algorithm II runs in time $O(e^2)$.

3. n-COLORABILITY OF COMPLETE GRAPHS

Although the problem of finding the n-colorability of arbitrary graphs is difficult we can say more for certain classes of graphs, e.g., c(G,n) = 1 if G is planar and n ≥ 4. For the class of complete graphs we can give an exhaustive a priori solution since all vertices are interchangeable.

Let π be a n-partition of $V(K_p)$. Then $B(K_p,\pi) = \sum_{i=1}^{n} \binom{m_i}{2}$ if π partitions $V(K_p)$ in (nonempty) sets of cardinality $m_1, m_2, \ldots, m_n$. Since $\sum_{i=1}^{n} \binom{m_i}{2} = 1/2(\sum_{i=1}^{n} m_i^2 - p)$ we have (if $\lfloor\ \rfloor$ and $\lceil\ \rceil$ denote the lower and upper entier, respectively).

<u>LEMMA</u> 4. $\inf_{\pi} B(K_p,\pi) = 1/2((n-r)\lfloor p/n \rfloor^2 + r\lceil p/n \rceil^2 - p)$
where r *is the remainder of* p/n.

<u>PROOF</u>. Let m be the smallest integer out of $m_1, m_2, \ldots, m_n$, $\sum_{i=1}^{n} m_i = p$. Then,

$$\sum_{i=1}^{n} m_i^2 = \sum_{i=1}^{n} (m + (m_i - m))^2 = \sum_{i=1}^{n} (m^2 + 2(m_i - m)m + (m_i - m)^2)$$

which is minimal for $(m_i - m) \in \{0,1\}$ for all i, $1 \le i \le n$. Hence a partition π_n of p into $m_1, m_2, \ldots, m_r = \lceil p/n \rceil$ and $m_{r+1}, m_{r+2}, \ldots, m_n = \lfloor p/n \rfloor$ yields a minimal $\sum_{i=1}^{n} m_i^2$. □

<u>COROLLARY</u> 5. $\inf_{\pi} B(K_p,\pi) = (p(p-n) - r(r-n))/2n$
(Substitute $(p/n - r/n)$ *for* $\lfloor p/n \rfloor$ *and* $(p/n + (n-r)/n)$ *for* $\lceil p/n \rceil$ *in Lemma 4.)*

<u>THEOREM</u> 6. $d(K_p,n) = \dfrac{(p+r-n)(p-r)}{np(p-1)}$
where r *is the remainder of* p/n.

Clearly, $\lim_{n\to\infty} d(K_p,n) = 1/n$ and, furthermore, for n = 2 $d(\cdot)$ is non-decreasing in p and assumes identical values for consecutive odd-even p's. In general:

<u>LEMMA</u> 7. *Let* $n \ge 2$ *be fixed.*
(i) $d(K_p,n) < \frac{1}{n}$ *for all* p.
(ii) $d(K_{\ell n+r},n) < d(K_{\ell n+r+1},n) < d(K_{\ell n+n-1},n) = d(K_{\ell n+n},n)$
where $0 \le r < n-1$.

PROOF. (i) By Theorem 6.

(ii) Assume that $d(K_{\ell n+r},n) \geq d(K_{\ell n+r+1},n)$ for some $n \geq 2$, $\ell \geq 1$ and r, $0 \leq r < n-1$. By Theorem 6 we have

$$\frac{(\ell n + 2r - n)\ell n}{n(\ell n + r)(\ell n + r - 1)} \geq \frac{(\ell n + 2r + 2 - n)\ell n}{n(\ell n + r + 1)(\ell n + r)}$$

which leads to $r \geq n-1$: contradiction. Similarly we prove the remaining inequalities/equalities. □

EXAMPLE. The Erdös-graph [2] has the mathematicians for its vertices; two vertices are joined by an edge if the corresponding mathematicians have co-authored at least one joint paper. What is the size of a possible clique in the Erdös graph in which the number of edges joining equal sexes is minimal, equals the number of edges joining opposite sexes and which clique contains 4% more female than male mathematicians? (The latter female predominance corresponds to nature's bias towards female births). From Theorem 6 we infer that the size of this clique is 625, viz. 325 females and 300 males. Using $\approx$ 3.8% instead of 4% yields the ominous clique K_{666}, and using $^1/_{600}$% allows the world population of 3.6 billion in such a clique.

4. EQUICOLORED SUBGRAPHS.

The number of equicolored K_m in K_p is, for a partition π of p in $m_1, m_2, \ldots, m_n$, equal to $\sum_{i=1}^{n} \binom{m_i}{m}$ (assuming $\binom{x}{y} = 0$ for $x < y$). The ratio of equicolored K_m to the total number of K_m in K_p is $\sum_{i=1}^{n} \binom{m_i}{m} / \binom{p}{m}$. The total number of equicolored complete graphs in a partition π of p in $m_1, m_2, \ldots, m_n$ is given by $\sum_{i=1}^{n} \sum_{j=1}^{m_i} \binom{m_i}{j}$ and the ratio to the total number of complete subgraphs is given by $\sum_{i=1}^{n} \sum_{j=1}^{m_i} \binom{m_i}{j} / \sum_{j=1}^{p} \binom{p}{j}$. Similarly, the number of equicolored subgraphs in K_p for a partition π is given by $\sum_{i=1}^{n} \sum_{j=1}^{m_i} f(j)\binom{m_i}{j}$ and the ratio by $\sum_{i=1}^{n} \sum_{j=1}^{m_i} f(j)\binom{m_i}{j} / \sum_{j=1}^{p} f(j)\binom{p}{j}$ where $f(j)$ is the number of graphs on j vertices.

LEMMA 8. *For all above expressions with a given* n *(and* m *if relevant) the infimum is reached for a partition of* $V(K_p)$ *in* $V_1, V_2, \ldots, V_n$ *of cardinality* $m_1, m_2, \ldots, m_n$ *where* $m_1, m_2, \ldots, m_r = \lceil \frac{p}{n} \rceil$ *and* $m_{r+1}, m_{r+2}, \ldots, m_n = \lfloor \frac{p}{n} \rfloor$ *where* r *is the remainder of* $\frac{p}{n}$.

PROOF. Let π be a partition of p in $m_1, m_2, \ldots, m_n$ and let $m' = \min\{m_1, m_2, \ldots, m_n\}$. Then, if $m_i = m'+j_i$ we have

$$\sum_{i=1}^{n} \binom{m'+j_i}{m} = \sum_{i=1}^{n} \frac{m'+j_i}{m'+j_i-m} \; \frac{m'+j_i-1}{m'+j_i-1-m} \cdots \frac{m'+1}{m'+1-m} \binom{m'}{m}$$

which is minimal for $j_i \in \{0,1\}$ for all i, $1 \le i \le n$, since $\frac{m'+x}{m'+x-m} > 1$ for all x. It is easy to see that all expressions alluded to in the Lemma reach their minimum if this particular one does so. □

The number of distinct labelled graphs over j vertices is found by noting that every edge in K_j can either be present or absent. Hence this number is $2^{\binom{j}{2}}-1$ (if we exclude the empty graph). The number of equicolored graphs in K_p for a partition π of p into $m_1, m_2, \ldots, m_n$ is therefore given by

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(2^{\binom{j}{2}}-1\right)\binom{m_i}{j},$$

and by Lemma 8 and the previous discussion we have

THEOREM 9. *The minimum number of equicolored graphs in an n-partition of* K_p *is given by*

$$r \sum_{j=1}^{\frac{p-r}{n}+1} \left(2^{\binom{j}{2}}-1\right)\binom{\frac{p-r}{n}+1}{j} + (n-r) \sum_{j=1}^{\frac{p-r}{n}} \left(2^{\binom{j}{2}}-1\right)\binom{\frac{p-r}{n}}{j}$$

Similarly, we determine the formulas for the infima of the other expressions occurring above.

ACKNOWLEDGEMENTS

I thank P. van Emde Boas, J.K. Lenstra, L. Meertens and D. de Werra for their comments.

REFERENCES

[1] EHRLICH, G., *Generator of set-partitions to exactly r subsets*, CACM 17 (1974) pp. 224-225.

[2] ERDÖS, P., *On the fundamental problem of mathematics*, Am. Math. Monthly 79 (1972) pp. 149-150.

[3] GAREY, M.R. & D.S. JOHNSON, *The complexity of near-optimal graph coloring*, JACM 23 (1976) pp. 43-49.

[4] HARARY, F., *Graph Theory*, Addison-Wesley (1969).

[5] JORDAN, K., *Calculus of Finite Differences*, Chelsea, New York (1967).

[6] KARP, R.M., *Reducibility among combinatorial problems*, in: (R.E. Miller and J.W. Thatcher, eds.) Complexity of Computer Computations, Plenum Press (1972).

[7] LENSTRA, J.K., *Sequencing by Enumerative Methods*, Ph.D. Thesis Univ. of Amsterdam (1976).

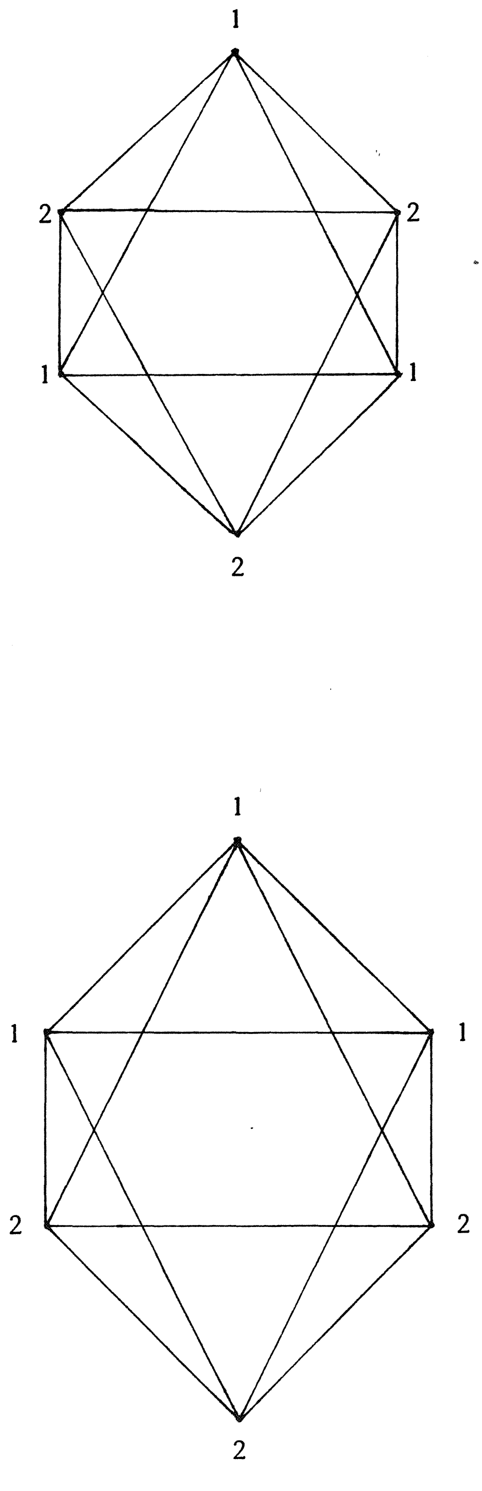

Fig. 1.