Annales Academiæ Scientiarum Fennicæ
Series A. I. Mathematica
Volumen 6, 1981, 135—148

# ON COLOR-FAMILIES OF GRAPHS

ARTO SALOMAA

**1. Introduction.** A mechanism for constructing families of "similar" graphs was introduced in [11]. The mechanism is essentially identical to the one considered in the theory of grammar forms and, at the same time, it generalizes the notion of coloring of graphs.

More specifically, one of the major trends in language theory during recent years has been the investigation of grammar forms. Starting from a "master grammar" one defines an "interpretation" mechanism giving rise to a family of grammars similar to the master grammar. (See [6]—[10], [14], [16], [18].) The relative position of language classes generated by such families of similar grammars has been one of the major concerns in grammar form theory. In particular, dense hierarchies of language families have been constructed by purely grammatical means, [8]—[10].

Such an interpretation mechanism can be defined in exactly the same way in a much more general set-up, for instance, for universal algebras. As far as we can see, the study of such a general case is of little interest only. However, the application of this mechanism to graphs (directed or undirected) has turned out to be of special interest. The reason for this is two-fold. (i) A classification of graphs is obtained, based on a notion that generalizes the notion of coloring in a natural way. (ii) Questions concerning this classification can be identified with questions concerning a fragment of the theory of grammar forms.

Point (ii) provides an interesting link between language and graph theory. When problems in one area can be transformed into problems in another area, it is likely that also techniques and results of the former area become applicable in the latter. Indeed, essentially graph-theoretic methods were applied in [13] to settle a wellknown open problem concerning the maximal density of a hierarchy of language families. It is likely that many similar "inter-disciplinary" applications can be given.

This paper deals only with graph-theoretic aspects of color-families; interconnections to the theory of grammar forms are mentioned very briefly. With the exception of a passage at the end of Section 4, considerations are restricted to undirected graphs. A brief outline of the contents of this paper follows.

Preliminary definitions and results are presented in Section 2. Section 3 deals with the density of color-families. The important result due to Welzl is established

doi:10.5186/aasfm.1981.0619

by a construction different from the one given in [17]. In Section 4, we consider color-families of infinite graphs. The main result is that if two graphs define the same color-family of finite graphs then they define also the same color-family of infinite graphs. This result holds for digraph families as well and, from the point of view of language theory, concerns languages over infinite alphabets, [1]. The final section contains scattered remarks, for instance, about the "master graph" being symmetric. Also some remarks about infinite master graphs are included.

The reader is assumed to be familiar with the basics of graph theory. For unexplained notions, [4] or [5] may be consulted. For the understanding of the technical details of this paper no knowledge of language theory is required on the part of the reader. A reader interested in the corresponding aspects of grammar form theory is referred first to [15] and then to [11]—[13] and [16]—[18], especially to [18].

**2. Preliminaries.** The graphs we consider have neither multiple edges nor loops (i.e., no edges from a vertex to itself). Such graphs are often referred to in the literature as "simple" graphs. Unless specified otherwise, our graphs will always be *undirected*. For a graph $G$, we denote by $V(G)$ (resp. $E(G)$) the set of vertices (resp. edges) of $G$. We consider both *finite* and *infinite* graphs, i.e., no over-all assumptions are made about the cardinality of $V(G)$, except that $V(G)$ is assumed to be nonempty.

Two vertices $x$ and $y$ in $V(G)$ are *adjacent* or *neighbours* if there is an edge between them. This fact is denoted by $A_G(x, y)$ or briefly by $A(x, y)$ if $G$ is understood. It is to be emphasized that no vertex is adjacent to itself (because the graphs we consider have no loops).

If $X$ is a nonempty subset of $V(G)$, we denote by $[G, X]=G_1$ the subgraph of $G$ generated by $X$. Thus,

$$V(G_1) = X \quad \text{and} \quad E(G_1) = E(G)|X.$$

(We use a vertical bar to denote restrictions of relations and functions in the customary fashion.)

For the *union* of two graphs, $G=G_1 \cup G_2$, we have

$$V(G) = V(G_1) \cup V(G_2) \quad \text{and} \quad E(G) = E(G_1) \cup E(G_2).$$

($E(G)$ is the set-theoretic union, and so no multiple edges will result.) Sometimes we use the union-sign also to denote the addition of some particular edge or edges to a given graph. Thus, we may write simply $G_1=G\cup(x, y)$, rather than specifying the second term of the union as a graph (which can of course always be done).

We now define some special graphs considered throughout this paper. For an integer $n\geqq 1$ (resp. $n\geqq 3$), we denote by $K_n$ (resp. $C_n$) the *complete graph* (resp. the *cycle*) with $n$ vertices. Thus, in $K_n$ there is an edge between any two distinct vertices, and in $C_n$ there are exactly $n$ edges such that the whole graph forms a cycle.

Observe that $K_3=C_3$ and that $K_1$ consists of one vertex and no edges. We consider also the graph $K_\infty$, where the vertices are in one-to-one correspondence with natural numbers, and there is an edge between any two distinct vertices.

From now on we consider in Sections 2 and 3 finite graphs only; all further definitions and discussions involving infinite graphs will be postponed until Section 4.

An *elementary morphism* in a graph $G$ consists of identifying two non-adjacent vertices $x$ and $y$ and inserting an edge between the identified vertex $x=y$ and all vertices $z$ adjacent to either $x$ or $y$ in $G$. A graph $G'$ is a *morphic image* of a graph $G$ if it is obtained from $G$ by finitely many elementary morphisms. $G$ is also considered to be a morphic image of itself. A graph $G$ is *minimal* if none of its morphic images, apart from $G$ itself, is a subgraph of $G$.

We now introduce the most important notions of this paper.

A graph $H$ is *colorable* according to a graph $G$, in symbols $H \leqq_c G$, if there is a mapping $\varphi$ of $V(H)$ into $V(G)$ such that, for all $x$ and $y$ in $V(H)$,

$$A_H(x,y) \quad \text{implies} \quad A_G(\varphi(x),\varphi(y)).$$

The mapping $\varphi$ is referred to as the *coloring* of $H$ according to $G$.

Assume that $G=K_n$, where $n\geqq 2$. Then clearly $H\leqq_c G$ if and only if $H$ is $n$-colorable in the customary sense. Hence, the notion defined above is natural extension of the customary notion of coloring of the vertices.

Every graph $G$ defines a family $\mathscr{L}(G)$ of graphs, consisting of all graphs colorable according to $G$. In symbols,

$$\mathscr{L}(G) = \{H \mid H \leqq_c G\}. \tag{1}$$

Families of graphs of the form (1) are referred to as *color-families*. (According to the convention made above, $G$ is a finite graph and every color-family consists of finite graphs. Infinite graphs will be considered in this sense in Section 4.)

Clearly, every family $\mathscr{L}(G)$ is infinite. No graph $G$ is universal in the sense that every finite graph would be contained in $\mathscr{L}(G)$. This follows because if $G$ has $n$ vertices, then the complete graph $K_{n+1}$ is not in $\mathscr{L}(G)$. The Four-Color Theorem tells us that every planar graph is in $\mathscr{L}(K_4)$.

As an example, the reader might want to consider the cyclic graph $C_5$. A graph $H$ is colorable according to $C_5$ if and only if it is 5-colorable in such a way that the adjacencies in $C_5$ are satisfied: if a vertex is colored by 1 then its neighbours may be colored by 2 or 5 but not by 3 or 4, and so forth.

The reader should have no difficulties in verifying the following inclusions:

$$\begin{gathered} \ldots \mathscr{L}(C_{2n+1}) \subsetneqq \mathscr{L}(C_{2n-1}) \subsetneqq \ldots \subsetneqq \mathscr{L}(C_5) \\ \subsetneqq \mathscr{L}(C_3) = \mathscr{L}(K_3) \subsetneqq \mathscr{L}(K_4) \subsetneqq \ldots \subsetneqq \mathscr{L}(K_n) \subsetneqq \mathscr{L}(K_{n+1}) \ldots . \end{gathered} \tag{2}$$

Indeed, (2) is referred to in [11] as the *basic* hierarchy of color-families. This reflects the fact that the generating graphs in (2) are very natural ones and, consequently, other color-families should be compared with the families (2).

Clearly, every family in the hierarchy (2) contains the family $\mathscr{L}(K_2)$. The family $\mathscr{L}(K_2)$ consists of all 2-colorable graphs, i.e., of all graphs having no cycles of an odd length. On the other hand, let $G$ be an arbitrary graph possessing a cycle of an odd length. Then there are $m$ and $n$ such that

$$\mathscr{L}(C_{2m+1}) \subseteqq \mathscr{L}(G) \subseteqq \mathscr{L}(K_n).$$

This is the best result we can get because it is a consequence of a result of Erdős, [3], that for any $m$ and $n$, there is a graph $G$ such that $\mathscr{L}(G)$ is incomparable with each of the families

$$\mathscr{L}(C_{2m-1}), \ldots, \mathscr{L}(C_3) = \mathscr{L}(K_3), \ldots, \mathscr{L}(K_{n-1}).$$

The reader is referred to [11] for further details.

The following results from [11] and [17] are preliminary ones and also not too difficult to deduce directly from the definitions.

Theorem 1. *The relation $\leqq_c$ is transitive. The inclusion $\mathscr{L}(H) \subseteqq \mathscr{L}(G)$ holds if and only if $H \leqq_c G$. Consequently, $\mathscr{L}(H) = \mathscr{L}(G)$ if and only if both $H \leqq_c G$ and $G \leqq_c H$. The relation $\leqq_c$ is decidable. Consequently, the relations*

$$\mathscr{L}(H) \subseteqq \mathscr{L}(G) \quad \text{and} \quad \mathscr{L}(H) = \mathscr{L}(G)$$

*are decidable for given graphs $H$ and $G$.*

Theorem 2. *A graph $H$ is $G$-colorable if and only if a morphic image $H_1$ of $H$ is isomorphic to a subgraph of $G$.*

Theorem 3. *For every color-family $\mathscr{L}$, there is a minimal graph $G$ such that $\mathscr{L} = \mathscr{L}(G)$. Moreover, if $G_1$ and $G_2$ are non-isomorphic minimal graphs, then the color-families $\mathscr{L}(G_1)$ and $\mathscr{L}(G_2)$ are different.*

Theorem 3 shows that minimal graphs constitute a suitable "normal form" of graphs for the representation of color-families.

Two graphs $H$ and $G$ are termed *color equivalent* if $\mathscr{L}(H) = \mathscr{L}(G)$. In accordance with Theorem 1 we denote this relation by $H =_c G$. Finally, the notation $H <_c G$ is used to mean that $H$ is colorable according to $G$ but $H$ and $G$ are not color equivalent. For instance, if $H$ and $G$ are cycles of an even length, the $H =_c G =_c K_2$. If $G$ is a cycle of an odd length and $H$ is a cycle of a greater odd length (or a cycle of an even length), then $H <_c G$.

**3. Density of color-families.** Consider the graphs $K_1$ and $K_2$. Clearly, $K_1 <_c K_2$ and, furthermore, there are no graphs $G$ satisfying

$$K_1 <_c G <_c K_2.$$

However, the pair $(K_1, K_2)$ constitutes the only example where one graph is a *predecessor* of another graph in the sense of the relation $<_c$. This result will be established in this section as the next theorem. We use the term "a nontrivial graph" to mean a graph possessing at least one edge.

Theorem 4. *Assume that $G_1$ and $G_2$ are nontrivial graphs satisfying $G_1 <_c G_2$. Then there is a graph $G_3$ such that*

$$G_1 <_c G_3 <_c G_2.$$

We shall first establish some lemmas. All graphs considered are assumed to be nontrivial. Following Welzl, [17], we say that $H$ is a *weak predecessor* of $G$ if $H <_c G$ and, moreover, apart from $H$ itself there is no morphic image $H_1$ of $H$ satisfying $H_1 <_c G$.

Thus, every odd cycle $C_{2m+1}$, $m \geqq 2$, is a weak predecessor of $C_{2m-1}$. The reader is encouraged to construct weak predecessors of $C_3$ of different types.

Lemma 4.1. *Assume that $H <_c G$. Then a weak predecessor $H$ of $G$ may be effectively obtained from $H$ by finitely many elementary morphisms. Moreover, if the vertices $x$ and $y$ of $H$ (and possibly some other vertices of $H$ as well) have been identified to form a single vertex of $H_1$ and if $\psi$ is the elementary morphism of $H$ identifying $x$ and $y$, then $\psi(H) <_c G$.*

*Proof.* The first assertion follows by Theorem 1. We just check through the vertices of $H$ whether or not there is an elementary morphism $\psi'$ such that $\psi'(H) <_c G$. If not, we choose $H_1 = H$. Otherwise, we repeat the procedure for an arbitrarily chosen $\psi'(H)$. (Observe that the process is nondeterministic: differently chosen elementary morphisms may give rise to different weak predecessors.)

The second assertion is due to the following observation. Assume that $\psi_1$ is the composition of elementary morphisms applied in constructing $H_1$ from $H$. Then $\psi_1$ can also be expressed as a composition where the first factor is $\psi$. □

Lemma 4.2. *Assume that a connected graph $G$ possesses, for every natural number $k$, a weak predecessor $H_k$ with more than $k$ vertices. Then $G$ has no predecessor.*

*Proof.* Assume the contrary: $G$ has a predecessor $P$. Hence, $P <_c G$ and there is no graph $P_1$ with the property

$$P <_c P_1 <_c G. \tag{3}$$

We assume without loss of generality (see Theorems 1 and 3) that $P$ and $G$ are minimal.

Let now $k$ be a fixed natural number greater than the cardinality of $V(P)$. Consider the weak predecessor $H_k$ of $G$ (as in the statement of Lemma 4.2).

Clearly, $P <_c H_k$ contradicts (3) with $P_1 = H_k$. Assume that

(4) $$H_k \leqq_c P.$$

This implies, by Theorem 2, that a morphic image $H_k'$ of $H_k$ is isomorphic to a subgraph of $P$. Consequently, $H_k' <_c G$. Since $H_k$ is a weak predecessor of $G$, this implies that $H_k' = H_k$. But this is impossible because, by the choice of $k$, $V(H_k)$ is of greater cardinality than $V(P)$. Hence, (4) cannot hold and, thus, $H_k$ and $P$ are incomparable with respect to the relation $\leqq_c$.

Consider the graph $H_k \cup P$. Clearly,

(5) $$P <_c H_k \cup P.$$

(For if $H_k \cup P$ is colorable according to $P$, then also (4) holds, which we have shown not to be the case.) By the definition of $H_k$ and $P$,

$$H_k \cup P \leqq_c G.$$

If $H_k \cup P <_c G$, then by (5) we obtain a contradiction by choosing in (3) $P_1 = H_k \cup P$. Consequently,

$$G \leqq_c H_k \cup P.$$

Since $G$ is connected and neither $G \leqq_c H_k$ nor $G \leqq_c P$ holds, we have a contradiction. (Observe that if a connected graph is colorable according to a union of graphs, then it must be colorable according to one of the components.) Consequently, $G$ has no predecessor. □

Lemma 4.3. *For all $m \geqq 1$ and $n \geqq 1$, the odd cycle $C_{2m+1}$ possesses a weak predecessor $D_{2m+1}^{2n+1}$ having at least $2n+1$ vertices.*

*Proof.* We first construct a graph $E_{2m+1}^{2n+1}$ as follows. Consider the cycle $C_{(2n+1)(2m+1)}$. Color its vertices, using $2n+1$ times the sequence of colors

$$1, 2, \ldots, 2m+1$$

in the clockwise order. To obtain $E_{2m+1}^{2n+1}$, add $(n-1)(2n+1)$ edges to $C_{(2n+1)(2m+1)}$ as follows. Each vertex colored with 1 is connected with an edge to the next (in the clockwise order) $n-1$ vertices colored with 2. (The vertex colored with 2 adjacent to the original vertex colored with 1 is not counted among these $n-1$ vertices. After all, it already is connected with an edge to the original vertex.)

The coloring given above shows that

$$E_{2m+1}^{2n+1} \leqq_c C_{2m+1}.$$

We want to show that, in fact,

(6) $$E_{2m+1}^{2n+1} <_c C_{2m+1}.$$

Clearly, (6) follows if the length of the shortest odd cycle $C$ in $E_{2m+1}^{2n+1}$ exceeds $2m+1$. This is true if $C$ contains a vertex colored with one of the numbers $3, \ldots, 2m+1$

because no edge added to $C_{(2n+1)(2m+1)}$ involves such a vertex. On the other hand, $C$ cannot consist only of vertices colored with 1 and 2 because then the length of $C$ would be even. Consequently, (6) holds.

Let now $D^{2n+1}_{2m+1}$ be a weak predecessor of $C_{2m+1}$, obtained from $E^{2n+1}_{2m+1}$ by the method of Lemma 4.1. We claim that no two vertices $x$ and $y$ of $E^{2n+1}_{2m+1}$ colored with 1 have been identified to form a single vertex of $D^{2n+1}_{2m+1}$. For if this would be the case then, by the second assertion in Lemma 4.1,

$$E' = \psi(E^{2n+1}_{2m+1}) <_c C_{2m+1}, \tag{7}$$

where $\psi$ is the elementary morphism of $E^{2n+1}_{2m+1}$ identifying $x$ and $y$. However, this is impossible because $E'$ has a cycle of length $2m+1$.

Consequently, the number of vertices in $D^{2n+1}_{2m+1}$ is not smaller than the number $2n+1$ of vertices colored with 1 in $E^{2m+1}_{2n+1}$. $\square$

We are now in the position to prove Theorem 4. Because $\mathscr{L}(K_2)$ consists of all graphs having no odd cycles, it suffices to prove that if $G$ is an arbitrary minimal graph with an odd cycle, then $G$ possesses no predecessor. Let $2m+1$, $m \geqq 1$, be the length of the shortest odd cycle in $G$.

Assume first that $G$ is connected. By Lemma 4.2, it suffices to prove that, for all $k$, $G$ possesses a weak predecessor $H_k$ with more than $k$ vertices. Consider a fixed number $k$ and choose $n$ in such a way that $2n+1>k$.

Consider a cycle $C$ in $G$ of length $2m+1$. Color its vertices with $1, 2, \ldots, 2m+1$. Then remove $C$ from $G$, replacing it with the graph $D^{2n+1}_{2m+1}$ defined in Lemma 4.3. For each $i$ such that $1 \leqq i \leqq 2m+1$ and each vertex $x$ colored with $i$ in $D^{2n+1}_{2m+1}$, construct an edge between $x$ and every vertex in $V(G-C)$ that was adjacent in the original $G$ to the vertex colored with $i$ in $C$. (Observe that because

$$D^{2n+1}_{2m+1} <_c C_{2m+1},$$

we may speak of the coloring of $D^{2n+1}_{2m+1}$ as above.) Denote by $G(n)$ the graph thus constructed.

It follows immediately from the construction that $G(n) \leqq_c G$. We want to show that

$$G(n) <_c G. \tag{8}$$

Assume the contrary: $G$ is colorable according to $G(n)$. Consequently, a morphic image $G'$ of $G$ appears as a subgraph in $G(n)$. Hence,

$$G' \leqq_c G(n) \leqq_c G.$$

Thus, a morphic image $G''$ of $G'$ (and hence also of $G$) is a subgraph of $G$. The minimality of $G$ now implies that

$$G'' = G' = G.$$

Therefore, $G$ is a subgraph of $G(n)$. The following argument shows that this is impossible and, consequently, (8) holds. Consider any occurrence of $G$ as a subgraph of $G(n)$. This occurrence cannot involve two vertices $x$ and $y$ of $D_{2m+1}^{2n+1}$ colored with the same color. For the identification of $x$ and $y$ would lead to a morphic image of $G$ that is also a subgraph of $G$, as seen by considering the identification in the whole $G(n)$. However, this contradicts the minimality of $G$.

Consequently, our occurrence of $G$ as a subgraph of $G(n)$ involves at most $2m+1$ vertices of $D_{2m+1}^{2n+1}$. On the other hand, it cannot involve less than $2m+1$ such vertices because then it would have altogether less vertices than the cardinality of $V(G)$. Hence, the occurrence of $G$ as a subgraph of $G(n)$ involves exactly $2m+1$ vertices of $D_{2m+1}^{2n+1}$, all colored differently. But this means that the occurrence misses at least one edge of $G$. This contradiction shows that (8) holds.

We now use the method of Lemma 4.1 to convert $G(n)$ into a weak predecessor $H_k$ of $G$. (Indeed, it can be shown that $H_k = G(n)$ but we do not need this information here.) Exactly as in the proof of Lemma 4.3 we see that in this process no two vertices of $D_{2m+1}^{2n+1}$ colored with 1 can be identified. Consequently, $H_k$ possesses at least $2n+1>k$ vertices, as required.

Assume, secondly, that $G$ is not connected. Thus,

$$G = G_1 \cup G_2 \cup \ldots \cup G_t \quad (t \geqq 2),$$

where each $G_i$ is connected. Furthermore, because of the minimality of $G$, the connected components $G_i$ are pairwise incomparable with respect to the relation $\leqq_c$.

Arguing indirectly, we assume that $P$ is a predecessor of $G$. Hence, $P <_c G$. This implies that one of the connected components of $G$, say $G_t$, is not colorable according to $P$. Consequently,

$$P \leqq_c P \cup G_1 \cup \ldots \cup G_{t-1} <_c G. \tag{9}$$

(Indeed, if $G$ would be colorable according to the union in (9), then also $G_t$ would be colorable according to this union. But this cannot be the case since $G_t$ is not colorable according to any of the components.)

If in (9) the sign $\leqq_c$ can be replaced by the sign $<_c$, we have contradicted the assumption that $P$ is a predecessor of $G$. Hence, we may assume that

$$P \cup G_1 \cup \ldots \cup G_{t-1} \leqq_c P.$$

Consequently, we obtain

$$P =_c P \cup G_1 \cup \ldots \cup G_{t-1} <_c G_1 \cup \ldots \cup G_{t-1} \cup G_t. \tag{10}$$

Let now $n$ be such that $2n+1$ exceeds the cardinality of $V(P)$. (Observe that $P$ need not be connected.) Construct $G_t(n)$ from $G_t$ in the same way as $G(n)$ in (8) was constructed from $G$.

We cannot have $G_t(n) \leqq_c P$ because this would imply that a morphic image of $G_t(n)$ is a subgraph of $P$. By the choice of $n$, this can happen only in case the morphic image in question is also a morphic image of $G_t$. But this cannot be the case because $G_t$ is not colorable according to $P$.

Exactly as in (8) we conclude that $G_t(n) <_c G_t$. (Observe that the minimality of $G$ yields the minimality of $G_t$.) These observations now yield the relations

$$P <_c P \cup G_t(n) <_c G, \tag{11}$$

contradicting the assumption that $P$ is a predecessor of $G$. (In fact, (10) is not needed to obtain (11). However, we wanted to give a detailed analysis about situations where the $G_t(n)$ graphs are not needed in the non-connected case.) We have, thus, completed the proof of Theorem 4 in all cases.

A few final remarks are in order. As pointed out already in the introduction Theorem 4 is due to Welzl, [17]. Our proof avoids the "super flowers" of [17] and is also different in other aspects. However, the crucial idea of using graphs $G(n)$ is due to Welzl. The proof of Theorem 4 is also constructive in the sense that it gives an algorithm for producing the graph $G_3$ strictly between the two given graphs $G_1$ and $G_2$. This can be seen by analysing the details of the argument.

**4. Color-families of infinite graphs.** We now extend the notion "colorable according to a graph $G$" to concern infinite graphs. The "master graph" $G$ will still be finite. In this way we obtain a method of characterizing families of infinite graphs. Such a method is of interest in view of the recent vivid discussion concerning constructive finitary specifications of infinite graphs. Moreover, because of interconnections with grammar forms referred to in the introduction, this approach is also linked with the recent study (see, for instance, [1]) concerning languages over infinite alphabets. We hope to return to this question in a forthcoming paper.

The reader is referred to Section 2 for our general conventions concerning graphs.

Let $G$ be a finite graph. A graph $H$ (finite or infinite) is *colorable according to* $G$, in symbols $H \leqq_c G$, if there is a mapping $\varphi$ of $V(H)$ into $V(G)$ such that, for all $x$ and $y$ in $V(H)$,

$$A_H(x, y) \quad \text{implies} \quad A_G(\varphi(x), \varphi(y)).$$

As before, we denote by $\mathscr{L}(G)$ the family of finite graphs colorable according to $G$. The notation $\mathscr{L}_\infty(G)$ stands for the family of all graphs, finite or infinite, colorable according to $G$. Also families of the latter type are referred to as *color-families.*

We shall prove in this section that, as far as color equivalence is concerned, the families $\mathscr{L}(G)$ are decisive. More specifically, we shall establish the following result.

Theorem 5. *For all finite graphs $G$ and $H$,*

$$\mathscr{L}_\infty(G) = \mathscr{L}_\infty(H) \quad \textit{if and only if} \quad \mathscr{L}(G) = \mathscr{L}(H).$$

The proof of Theorem 5 is based on the following "infinity lemma".

Lemma 5.1. *Assume that $G$ is a finite graph and that $H$ is an infinite graph such that all finite subgraphs of $H$ are in $\mathscr{L}(G)$. Then $H$ itself is in $\mathscr{L}_\infty(G)$.*

*Proof.* We assume without loss of generality that $G$ is minimal. For an arbitrary graph $D$, we say that $D$ *possesses the property* $P_G$, or shortly that $P_G(D)$ holds, if every finite subgraph of $D$ is in $\mathscr{L}(G)$, i.e., is colorable according to $G$. By the assumption, $P_G(H)$ holds.

Consider the (disjoint) union $G \cup H$. Clearly, $P_G(G \cup H)$ holds.

We now add edges to $G \cup H$ in such a way that the resulting graph still always possesses the property $P_G$. Let $T$ be a maximal graph obtained in this fashion. ($T$ need not be unique: the order in which the edges are added may affect it.) Thus, $T$ satisfies the following conditions (i)—(iii):

(i) $G \cup H$ is a subgraph of $T$ and $V(T) = V(G \cup H)$.

(ii) $P_G(T)$ holds.

(iii) Whenever an edge $(x, y)$ is missing from $T$, then the graph $T \cup (x, y)$ does not possess the property $P_G$, i.e., there is a finite subgraph of $T \cup (x, y)$ not belonging to $\mathscr{L}(G)$.

We shall prove that $T$ belongs to $\mathscr{L}_\infty(G)$. By (i), this implies that $H$ is in $\mathscr{L}_\infty(G)$.

In what follows we consider the relation $A_T$, written briefly $A$. Thus, $A(x, y)$ means that $x$ and $y$ are adjacent in $T$. We also use the notation $NA(x, y)$, meaning that $x$ and $y$ are not adjacent in $T$.

Consider also the following binary relation $\equiv$, defined on the set $V(T)$. By definition, $x \equiv y$ holds if and only if $x$ and $y$ are not adjacent in $T$ but possess the same neighbours in $T$. In other words, $x \equiv y$ holds if and only if $NA(x, y)$ holds and there is no $z$ such that exactly one of the relations $A(x, z)$ and $A(y, z)$ holds.

To prove that $\equiv$ is an equivalence, we first note that it clearly is both reflexive and symmetric. To establish transitivity, assume that $x \equiv y$ and $y \equiv z$. Then $NA(x, z)$ because, otherwise, we have both $A(x, z)$ and $NA(y, z)$, contradicting $x \equiv y$. If there is a $u$ such that $A(x, u)$ and $NA(z, u)$, then we have $A(y, u)$, contradicting $y \equiv z$. A similar contradiction arises from the assumptions $NA(x, u)$ and $A(z, u)$. Consequently, $x \equiv z$. This shows that the relation $\equiv$ is an equivalence.

Observe that intuitively $x \equiv y$ means that removing one of the vertices $x$ and $y$ gives the same result as the elementary morphism identifying $x$ and $y$. Before continuing the proof of Lemma 5.1, we establish the following result. Apart from the present proof, the result is useful also in many analogous situations.

Lemma 5.2. *Assume that $M$ is a minimal finite graph and $P$ a finite graph containing an occurrence of $M$ as a subgraph. Assume further that $P \equiv_c M$. Then there is a coloring $\alpha$ of $P$ according to $M$ such that $\alpha|M$ is the identity. (Here $\alpha|M$ denotes the restriction of $\alpha$ to the occurrence of $M$ in $P$ we are considering.)*

*Proof.* Consider an arbitrary coloring

$$\beta\colon V(P) \to V(M).$$

Then the restriction $\beta_1 = \beta|M$ is a coloring of $M$ according to $M$. Because of the minimality of $M$, $\beta_1$ is a permutation of $V(M)$.

If $\beta_1$ is the identity permutation, there is nothing to prove. Otherwise, we let $k>1$ be the order of $\beta_1$ (i.e., $\beta_1^k$ is the identity) and consider

$$\alpha = \beta^k.$$

Clearly, $\alpha|M$ is the identity. Because $\beta$ is a coloring of $P$ according to $M$ and because $M$ is a subgraph of $P$, it can be immediately verified that $\alpha$ is a coloring of $P$ according to $M$. □

We now return to the proof of Lemma 5.1. Consider the subgraphs

$$G_1 = [T, V(G)] \quad \text{and} \quad H_1 = [T, V(H)].$$

The minimality of $G$ implies that no two vertices of $G_1$ belong to the same equivalence class according to the relation $\equiv$.

We claim that, for every $x$ in $V(H_1)$, there is a $y$ in $V(G_1)$ such that $x \equiv y$. This claim immediately gives a coloring of $T$ according to $G$ and, consequently, shows the correctness of Lemma 5.1.

We shall establish our claim by an indirect argument. Thus, assume there is a vertex $x$ in $V(H_1)$ such that no vertex $y$ in $V(G_1)$ satisfies $x \equiv y$.

Let $y_1, \ldots, y_k$ be all the vertices of $G_1$ that are not adjacent to $x$ (in $T$). We observe first that there must be such vertices, i.e., $k \geqq 1$. For if $x$ is adjacent to every vertex in $V(G_1)$, we obtain a contradiction as follows. Consider the subgraph $[T, V(G_1) \cup \{x\}]$. Because $T$ possesses the property $P_G$, this subgraph is colorable according to $G$ or, equivalently, according to $G_1$. By Lemma 5.2, a coloring $\alpha$ can be chosen such that $\alpha$ is the identity on $V(G_1)$. But now $\alpha(x)$ cannot be defined because there are no loops in $G_1$. Thus, we must have $k \geqq 1$.

Since no vertex of $G_1$ is equivalent to $x$ (in the sense of $\equiv$), this applies also to the vertices $y_1, \ldots, y_k$. Consequently, for each $i=1, \ldots, k$, there is a vertex $x_i$ in $T$ such exactly one of the adjacencies

$$A(y_i, x_i) \quad \text{and} \quad A(x, x_i)$$

holds in $T$. For $i=1, \ldots, k$, let $z_i$ be the one among the vertices $y_i$ and $x$ that satisfies $NA(z_i, x_i)$.

By property (iii) of the graph $T$, if the edge $(z_i, x_i)$ is added to $T$, then the resulting graph does not possess the property $P_G$. Thus, for $i=1, \ldots, k$, there is a finite subgraph $D_i$ of $T$ such that the graph

$$D_i \cup (z_i, x_i) \tag{12}$$

is not colorable according to $G$. Hence, $z_i$ and $x_i$ are in $V(D_i)$.

Consider the subgraph

$$D = [T, V(G) \cup \{x\} \cup V(D_1) \cup \ldots \cup V(D_k)].$$

Since $D$ is finite, it is colorable according to $G$. By Lemma 5.2, $D$ possesses a coloring $\alpha$ (according to $G$) such that $\alpha|V(G)$ is the identity.

For $i=1, \ldots, k$, the mapping $\alpha | V(D_i)$ is a coloring of $D_i$ according to $G$. This implies that we have

(13) $$NA(\alpha(z_i), \alpha(x_i)) \quad (1 \leqq i \leqq k)$$

because, otherwise, the graph (12) would possess a coloring according to $G$ (namely, $\alpha$ restricted to (12)).

Because $\alpha | V(G)$ is the identity, there is a $j$, $1 \leqq j \leqq k$, such that

$$\alpha(x) = y_j .$$

Two cases arise.

Assume first that $z_j = x$. By the choice of $z_j$, we have $NA(x, x_j)$ and $A(y_j, x_j)$. On the other hand, by (13), we have $NA(\alpha(x), \alpha(x_j))$ and, hence, $NA(y_j, \alpha(x_j))$. From $A(y_j, x_j)$ we infer (because $\alpha(y_j) = y_j$ and $\alpha$ is a coloring) the relation $A(y_j, \alpha(x_j))$, which is a contradiction.

Assume, secondly, that $z_j = y_j$. This means that we have both $NA(y_j, x_j)$ and $A(x, x_j)$. On the other hand, (13) gives us $NA(\alpha(y_j), \alpha(x_j))$ and, hence, $NA(y_j, \alpha(x_j))$. But now $A(x, x_j)$ yields the relation $A(\alpha(x), \alpha(x_j))$, from which we obtain the contradictory relation $A(y_j, \alpha(x_j))$. This concludes the proof of Lemma 5.1.

Lemma 5.1 is also a generalization of the old result of [2], dealing with the customary notion of coloring. The use of the relation $\equiv$ introduced above simplifies also to some extent the proof given in [5] for the result of [2].

According to Lemma 5.1, an infinite graph $H$ is in $\mathscr{L}_\infty(G)$ if and only if all finite subgraphs of $H$ are in $\mathscr{L}(G)$. Since clearly the equation $\mathscr{L}_\infty(G) = \mathscr{L}_\infty(H)$ implies the equation $\mathscr{L}(G) = \mathscr{L}(H)$, we obtain now also Theorem 5.

We have considered in this paper only undirected graphs. Color-families can be defined for digraphs as well. (This is even more closely linked with language theory, because undirected graphs correspond to subsets of the free commutative monoid.)

We assume that the digraphs considered have neither loops nor multiple arrows. On the other hand, if $x$ and $y$ are two vertices of a digraph, it is possible that the digraph has an arrow from $x$ to $y$ and also an arrow from $y$ to $x$.

We say that a digraph $H$ (finite or infinite) is *colorable* according to a finite digraph $G$ if there is a mapping

$$\varphi\colon V(H) \to V(G)$$

such that, whenever there is an arrow from a vertex $x$ to a vertex $y$ in $H$, then there is also an arrow from $\varphi(x)$ to $\varphi(y)$ in $G$.

The families $\mathscr{L}(G)$ and $\mathscr{L}_\infty(G)$ are defined in the same way as before. Most of the results concerning graphs carry over to digraphs. However, nontrivial examples of predecessors can be given and, as regards density in general, the situation is more complicated for digraphs. The reader is referred to [11]—[13] for further details.

In particular, Lemma 5.1 and Theorem 5 can be extended to concern digraphs. This will be summarized in the following theorem. We omit the proof because it is analogous to the proofs given above. In particular, observe that the relation $\equiv$ (modified to digraphs in an obvious fashion) will still be an equivalence.

Theorem 6. *Assume that $G$ is a finite digraph and that $H$ is an infinite digraph such that all finite subdigraphs of $H$ are in $\mathscr{L}(G)$. Then also $H$ itself is in $\mathscr{L}_\infty(G)$. For all finite digraphs $G$ and $H$,*

$$\mathscr{L}_\infty(G) = \mathscr{L}_\infty(H) \quad \textit{if and only if} \quad \mathscr{L}(G) = \mathscr{L}(H).$$

**5. Concluding remarks.** We mention here briefly some topics and problems not discussed more closely in this paper.

Very little is known about color-families of particular graphs, such as the family $\mathscr{L}(C_5)$. Although, by Theorem 1, the membership problem is always decidable, the decision is in most cases hard. For instance, the membership in $\mathscr{L}(C_5)$ is an $NP$-complete problem (I. H. Sudborough, personal communication).

As regards Theorem 5 and Lemma 5.1, the consideration of subgraphs of some bounded size is not sufficient. More specifically, one can establish the following result. For every $k \geqq 2$ and every $n$, there is a minimal graph $G$ with $k$ vertices and an infinite graph $H$ such that all subgraphs of $H$ with at most $n$ vertices are colorable according to $G$ but $H$ itself is not colorable according to $G$.

In our considerations the "master graph" has always been finite, i.e., we have considered coloring according to a finite graph $G$ only. The definitions can be extended in an obvious fashion to the case where $G$ is infinite. Then there will be also "universal" graphs, for instance $K_\infty$. This is an immediate consequence of the fact that $\mathscr{L}(K_\infty)$ (resp. $\mathscr{L}_\infty(K_\infty)$) contains all finite (resp. denumerably infinite) graphs.

Perhaps the most interesting area of open problems will be the study of subcollections of color-families. The customary notion of coloring deals only with families $\mathscr{L}(K_i)$, where $K_i$ is a complete graph. We have considered in this paper families $\mathscr{L}(G)$, where $G$ is an arbitrary finite graph (or, equivalently, an arbitrary minimal graph).

Such a natural subcollection is obtained by considering "master graphs" $G$ with a transitive automorphism group. This means that the coloring according to such a $G$ is *symmetric* in the following sense. Assume that $H \leqq_c G$, and that $x$ in $V(H)$ and $y$ in $V(G)$ are arbitrary. Then there is a coloring $\varphi$ of $H$ according to $G$ such that $\varphi(x) = y$.

Clearly, all graphs $C_{2m+1}$ and $K_i$ have this property, whereas the graph $D_{2m+1}^{2n+1}$ discussed in Lemma 4.3 does not, in general, have this property. Clearly, graphs symmetric in this sense are $k$-regular for some $k$, whereas the converse is not true. We hope to return in a forthcoming paper to the characterization of this subcollection of color-families.

*Acknowledgement.* The author is indebted to H. A. Maurer, E. Welzl and D. Wood for useful discussions.

## References


[1] AUTEBERT, J.-M., J. BEAUQUIER, and L. BOASSON: Langages sur des alphabets infinis. - Discrete Appl. Math. 2, 1980, 1—20.

[2] BRUIJN, N. G. DE, and P. ERDÖS: A colour problem for infinite graphs and a problem in the theory of relations. - Nederl. Akad. Wetensch. Indag. Math. 13, 1951, 371—373.

[3] ERDÖS, P.: Graph theory and probability II. - Canad. J. Math. 13, 1961, 346—352.

[4] HARARY, F.: Graph theory. - Addison-Wesley Publishing Company, Reading, Massachusetts—Menlo Park, California—London—Don Mills, Ontario, 1972.

[5] LOVÁSZ, L.: Combinatorial problems and exercises. - North-Holland, 1979.

[6] MAURER, H. A., A. SALOMAA, and D. WOOD: EOL forms. - Acta Inform. 8, 1977, 75—96.

[7] MAURER, H. A., A. SALOMAA, and D. WOOD: Context-free grammar forms with strict interpretations. - J. Comput. System Sci. (to appear).

[8] MAURER, H. A., A. SALOMAA, and D. WOOD: Dense hierarchies of grammatical families. - J. Assoc. Comput. Mach. (to appear).

[9] MAURER, H. A., A. SALOMAA, and D. WOOD: MSW spaces. - Inform. and Control (to appear).

[10] MAURER, H. A., A. SALOMAA, and D. WOOD: Decidability and density in two-symbol grammar forms. - MCMASTER University Computer Science Technical Report 79—CS—20, 1979, submitted for publication.

[11] MAURER, H. A., A. SALOMAA, and D. WOOD: Colorings and interpretations — a connection between graphs and grammar forms. - McMaster University Computer Science Technical Report 80—CS—10, 1980, submitted for publication.

[12] MAURER, H. A., A. SALOMAA, and D. WOOD: On finite grammar forms. - McMaster University Computer Science Technical Report 80—CS—9, 1980, submitted for publication.

[13] MAURER, H. A., A. SALOMAA, and D. WOOD: On predecessors of finite languages. - McMaster University Computer Science Technical Report 80—CS—14, 1980, submitted for publication.

[14] ROZENBERG, G., and A. SALOMAA: The mathematical theory of L systems. - Academic Press, New York—London—Toronto—Sydney—San Francisco, 1980.

[15] SALOMAA, A.: Formale Sprachen. - Springer-Verlag, Berlin—Heidelberg—New York, 1978.

[16] SALOMAA, A.: Grammatical families. - Springer Lecture Notes in Computer Science 85, 1980, 543—554.

[17] WELZL, E.: Color-families are dense. - Theoret. Comput. Sci. (to appear).

[18] WOOD, D.: Grammar and L forms: An introduction. - Springer Lecture Notes in Computer Science 91, 1980.



University of Turku
Department of Mathematics
SF-20500 Turku 50
Finland


Received 17 November 1980