Discrete Applied Mathematics 3 (1981) 301–312
North-Holland Publishing Company

COMPLEXITY OF REPRESENTATION OF GRAPHS BY SET SYSTEMS

Svatopluk POLJAK, Vojtěch RÖDL and Daniel TURZÍK
KZAA, MFFUK, Sokolovská 83, 180 000 Praha 8, Czechoslovakia

Received 16 October 1980
Revised 25 March 1981

Let $\mathcal{F}$ be a family of subsets of S and let G be a graph with vertex set $V=\{x_A \mid A\in\mathcal{F}\}$ such that: (x_A, x_B) is an edge iff $A\cap B\neq\emptyset$. The family $\mathcal{F}$ is called a set representation of the graph G.

It is proved that the problem of finding minimum k such that G can be represented by a family of sets of cardinality at most k is NP-complete. Moreover, it is NP-complete to decide whether a graph can be represented by a family of distinct 3-element sets.

The set representations of random graphs are also considered.

1. Introduction

Let $G=(V,E)$ be a graph without loops and multiple edges. A family $\mathcal{F}=(A_x \mid x\in V)$ of (not necessarily distinct) sets is called a *set representation* of G if

$$A_x\cap A_y\neq\emptyset \quad \text{iff} \quad (x,y)\in E$$

for every pair x,y of distinct vertices of G; conversely G is called an *intersection graph* of $\mathcal{F}$. A set representation $\mathcal{F}$ of G is called a *k-set representation* if $|A_x|\leq k$ for all $x\in V$; a *distinct set representation* if $A_x\neq A_y$ for all $x,y\in V$, $x\neq y$, a *simple set representation* if $|A_x\cap A_y|\leq 1$ for all $x,y\in V$, $x\neq y$.

It is well known (see [12]) that every graph has a simple set representation.

We shall deal with the problems of finding optimal set representations for graphs under two optimization criteria:

(1) minimize the maximum size of the sets,

(2) minimize the size of the universe of elements.

The first criterion generalizes the question of line graphs because line graphs are the graphs with a distinct 2-set representation. Similarly, graphs with a 2-set representation are intersection graphs of multigraphs. Both these classes have a good characterization given in terms of finite number of forbidden subgraphs (see [1, 3]). These characterizations assert the existence of a polynomial time algorithm for determining whether a given graph has a (simple) 2-set representation. The number given by the criterion (2) is called the intersection number; it belongs to the long studied combinatorial quantities (see [5]) and is known to be NP-complete (see [15]). For special classes of graphs the intersection number is either given by a formula or is computable in polynomial time (see [9, 14]).

There are also interesting questions concerning set representations by families of special types, for example interval graphs (see [8]), intersection graphs of curves in the plane (see [4]), etc., but we shall not deal with these.

In Section 1 we transform the questions of set representation to the questions of covering by complete subgraphs, which is a more convenient approach.

In Section 2 we show that it is NP-complete to find a minimum integer k for which a given graph G has a k-set representation. It is even NP-complete to decide whether a given graph G has a 4-set representation. Moreover, it is NP-complete to decide whether a graph has a distinct 3-set representation. These results indicate that the characterization of line graphs probably cannot be generalized even for triples.

Further, in Section 3 we show that it is NP-complete to find the minimum k such that for a given graph G there exists a simple set representation with $|\bigcup \mathcal{F}| = k$. This result can also be considered in connection with line graphs because if G is a graph and $H = L(G)$, *its line graph, then* G is a simple set representation of H.

In Section 4 we discuss the structure of the set Forb_3 which is defined to be the set of minimal forbidden subgraphs for the class of graphs with 3-set representation.

In Section 5 we give some estimations for set representations of random graphs.

For the graph-theoretic terms used see [2], for details of reduction techniques see [10].

1. Covering of graphs

Let $G = (V, E)$ be a graph. A system $\mathfrak{A}$ of complete subgraphs of G is called a *cover* of G if every edge of G belongs to at least one complete graph from $\mathfrak{A}$. We say that a cover $\mathfrak{A}$ is: *k-cover* if every vertex of G belongs to at most k graphs from $\mathfrak{A}$, *edge-disjoint* if no two graphs from $\mathfrak{A}$ have a common edge, *vertex-separating* if for every pair of vertices of G there is a member of $\mathfrak{A}$ containing just one of them.

The following theorem (see [2]) gives a correspondence between set representations and covers of graphs and will be used implicitely.

Theorem 1.1. *Let $G = (V, E)$ be a graph. The following two mappings*

$$\mathcal{F} = (A_v \mid v \in V) \mapsto \mathfrak{A} = \{K_x \mid x \in U\} \quad \text{where } K_x = \{v \in V \mid x \in A_v\}, \quad U = \bigcup_{v \in V} A_v,$$

$$\mathfrak{A} = \{K_x \mid x \in U\} \mapsto \mathcal{F} = (A_v \mid v \in V) \quad \text{where } A_v = \{x \in U \mid v \in K_x\}$$

give a one-one correspondence between set representations ($\mathcal{F}$) and covers ($\mathfrak{A}$) of G. Moreover, a set representation $\mathcal{F}$ and its corresponding cover $\mathfrak{A}$ satisfy:

(i) $\mathcal{F}$ *is a k-set representation iff* $\mathfrak{A}$ *is a k-cover,*
(ii) $\mathcal{F}$ *is distinct iff* $\mathfrak{A}$ *is vertex separating,*
(iii) $\mathcal{F}$ *is simple iff* $\mathfrak{A}$ *is edge disjoint.*

2. Set representations with minimum size of sets

For a given graph G denote by $\tau(G)$ and $\tau_d(G)$ the minimum k for which there exists a k-cover and a distinct k-cover, respectively.

Theorem 2.1. *For a given graph G and an integer k it is NP-complete to decide whether* $\tau(G) \le k$.

Proof. For a given graph G with n vertices we shall construct a graph H such that

$$\chi(G) = \tau(H) - n, \tag{1}$$

where χ is the chromatic number. This reduces the determination of the chromatic number, which is NP-complete (see [11]), to the determination of τ. The graph H is constructed as follows: To the graph $\bar{G}$, the complement of G, add new vertices $y_1, y_2, \ldots, y_n, x$ and join the vertex x to all vertices of $V(G) \cup \{y_1, y_2, \ldots, y_n\}$. Consider a cover $\mathfrak{A}$ of H. Clearly

$$|\{K \mid x \in K \in \mathfrak{A}\}| \ge n + \chi(G)$$

which gives

$$\chi(G) \le \tau(H) - n.$$

On the other hand, suppose that G is colored by $\chi(G)$ colors and take a cover $\mathfrak{A}$ formed by following sets

$$\{x\} \cup \{v \mid v \in V(G),\ v \text{ is colored by } i\}, \quad i = 1, 2, \ldots, \chi(G),$$

$$\{u, v\} \quad \text{for all pairs } u, v \in V(G),\ (u, v) \notin E(G),$$

$$\{x, y_i\}, \quad i = 1, 2, \ldots, n.$$

Hence

$$\chi(G) \ge \tau(H) - n.$$

The *satisfiability problem* of Boolean expressions in conjunctive normal form with at most three literals per clause will be abbreviated by 3-SAT. The 3-SAT problem is known to be NP-complete (see [11]). We will consider the version of 3-SAT with exactly 3-distinct literals per clause (see e.g. [13]).

Theorem 2.2. *It is NP-complete to decide whether a given graph G has a distinct* 3-*set representation.*

Proof. We shall reduce to it the 3-SAT problem. Let $\Phi = c_1 \wedge \ldots \wedge c_m$ be a Boolean expression of variables $x, y, z, \ldots$ which is an instance of 3-SAT. We shall construct, in the following four steps, a graph G such that Φ is satisfiable iff there exists a vertex-separating 3-cover of G.

(1) For every variable x let H_x denote the graph given by Fig. 1 with some edges labeled by symbols $x_1, x_2, \ldots, x_m, \bar{x}_1, \ldots, \bar{x}_m$.

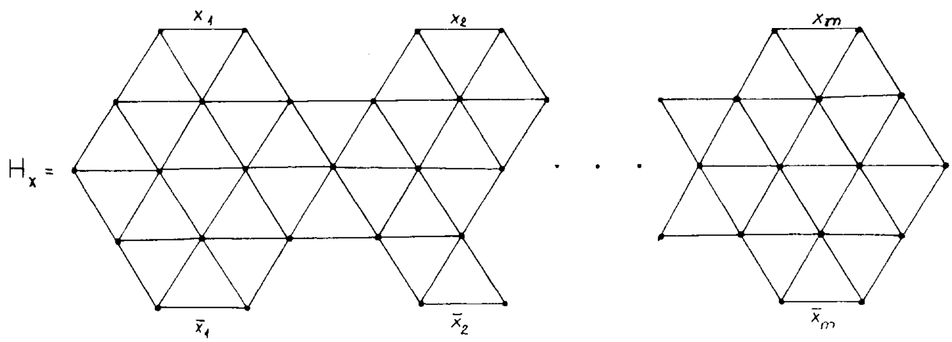

Fig. 1.

(2) For every $i = 1, 2, \ldots, m$ let us consider a graph F_i given by Fig. 2 with labeled vertices a_i, b_i and edges $\alpha_i, \beta_i, \gamma_i$, where α, β, γ are the literals appearing in the clause c_i.

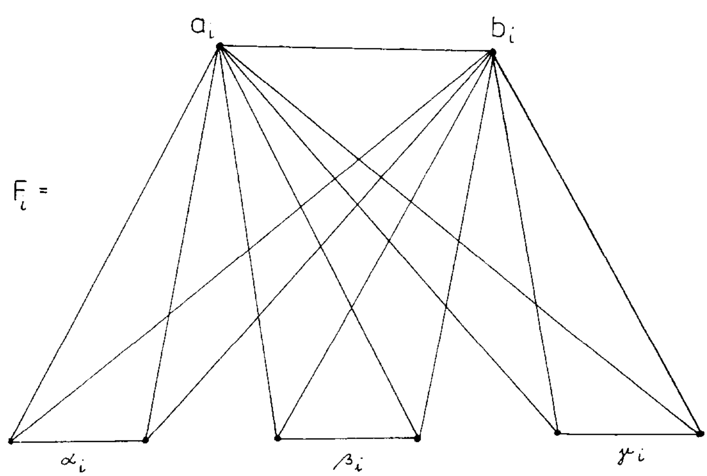

Fig. 2.

(3) Put $H = \sum H_x$, where x runs over all variables of Φ, and

$$F = \sum F_i, \quad i = 1, 2, \ldots, m.$$

(4) Let G be an amalgamation of F and H given by glueing edges with the same label.

The graphs used in the construction have the following properties:

(i) Every 3-cover $\mathfrak{A}_x$ of H_x satisfies: either all the edges $x_1, \dots, x_m$ are covered by triangles of $\mathfrak{A}_x$ and none of the $\bar{x}_i$ are, or all of the $\bar{x}_i$ are and none of the x_i.

(ii) A 3-cover $\mathfrak{A}_i$ of F_i separates vertices a_i, b_i iff at least one of edges $\alpha_i, \beta_i, \gamma_i$ is covered by two triangles of $\mathfrak{A}_i$.

Let $\mathfrak{A}$ be a vertex separating 3-cover of G. Let us consider a truth assignment t for Φ given by

$$(*) \qquad t(x)=0 \quad \text{iff (every) } x_i \text{ is covered by a triangle of } \mathfrak{A} \text{ in } H_x.$$

If α_i is covered by a triangle of $\mathfrak{A}$ in G_x then α_i must be covered by K_4 in F_i. Hence, using (ii), t is a satisfying truth assignment for Φ.

On the other hand, suppose that $t: \{x, y, z, \dots\} \to \{0, 1\}$ is a truth assignment satisfying Φ. Let us consider a cover $\mathfrak{A}$ of G consisting of

(a) the 3-cover of H_x satisfying (*) (for every variable x);

(b) all copies of K_4 in F_i containing α_i with $t(\alpha_i)=0$, and all triangles in F_i containing α_i with $t(\alpha_i)=1$ $(i=1, \dots, m)$.

Thus, by property (i), the 3-cover $\mathfrak{A}$ is vertex-separating.

Theorem 2.3. *It is NP-complete to decide whether a given graph G has a 4-set representation.*

Proof. We shall reduce to it the 3-SAT problem. Let Φ be an instance of 3-SAT as in the proof of the Theorem 2.2. We shall construct a graph G such that Φ is satisfiable iff there exists a 4-cover of G.

(1) For every variable x of Φ let H_x be a graph arising from the graph given by Fig. 1. after adding one pendant edge to every vertex of H_x.

(2) For every $i=1, \dots, m$ let us construct a graph F_i (with 16 vertices) in the following way. Consider three copies W_1, W_2, W_3 of 8-wheel given by Fig. 3 and identify these vertices: v_1 with v_1', v_2 with v_2' and v_2'', v_3 with v_3' and v_3'', v_4 with v_4' and v_4'', v_5' with v_5''.

(3) Put $H = \sum H_x$, where x runs over all variables of Φ, and

$$F = \sum F_i, \quad i=1, \dots, m.$$

(4) Let G be an amalgamation of H and F given by glueing edges with the same label.

The graphs used in the construction have the following properties.

(i) Every 4-cover $\mathfrak{A}_x$ of H_x satisfies: either all the edges $x_1, \dots, x_m$ are covered by triangles of $\mathfrak{A}_x$ and none of the $\bar{x}_i$ are, or all the $\bar{x}_i$ are and none of the x_i.

(ii) In every 4-cover $\mathfrak{A}_i$ of F_i there is at least one of edges $\alpha_i, \beta_i, \gamma_i$ covered by a triangle of $\mathfrak{A}_i$. (To see it consider the neighbourhood of the vertex $v_2 = v_2' = v_2''$.)

(iii) For each of $\alpha_i, \beta_i, \gamma_i$ there exists a 4-cover of F_i such that this edge is covered by a triangle and the other two are not (Fig. 4).

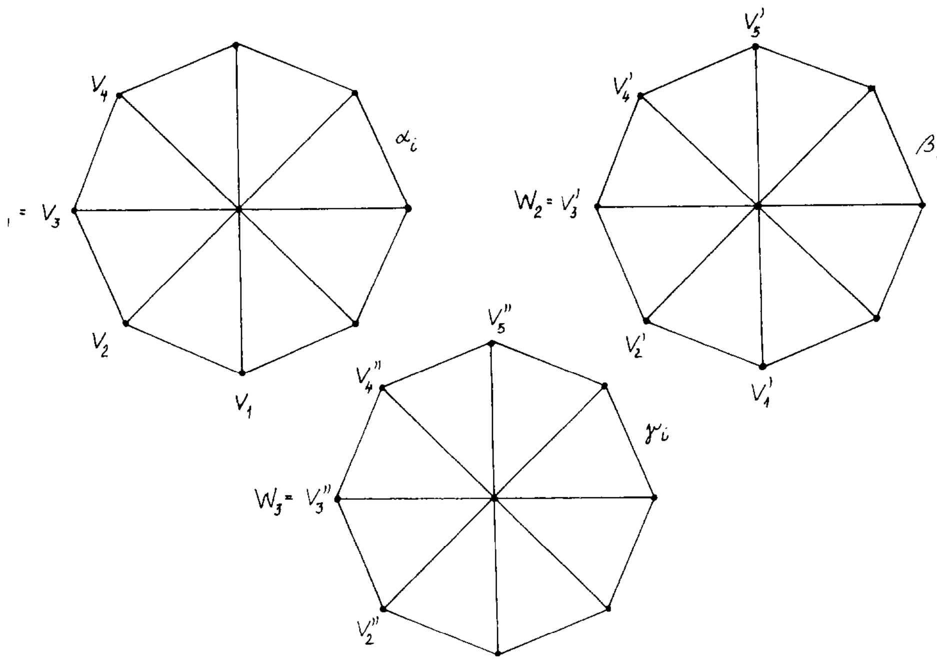
V4
αi
V3
V2
V1
V5'
V4'
βi
W2 = V3'
V2'
V1'
V5''
V4''
γi
W3 = V3''
V2''

Fig. 3.

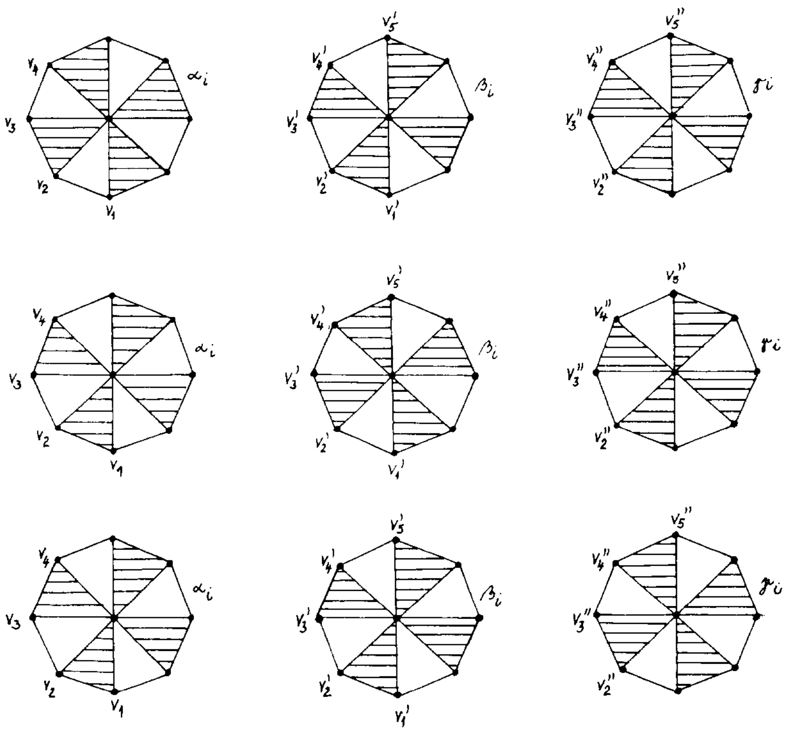
V4
αi
V3
V2
V1
V5'
V4'
βi
V3'
V2'
V1'
V5''
V4''
γi
V3''
V2''

Fig. 4.

Let $\mathfrak{A}$ be a 4-cover of G. Let us consider a truth assignment t for Φ given by (*). It follows from (ii) that t is a satisfying truth assignment for Φ. Conversely, if t is a satisfying truth assignment for Φ, then (iii) guarantees the existence of a 4-cover of G satisfying (*).

Theorem 2.4. *It is NP-complete to decide whether a given graph G has a simple* 3-*set representation.*

Proof. Modify the proof of the Theorem 2.2. as follows. Consider 3-covers which are edge-disjoint instead of vertex-separating. Let the graphs F_i be given by Fig. 5. Now, the proof runs as the proof of the Theorem 2.2.

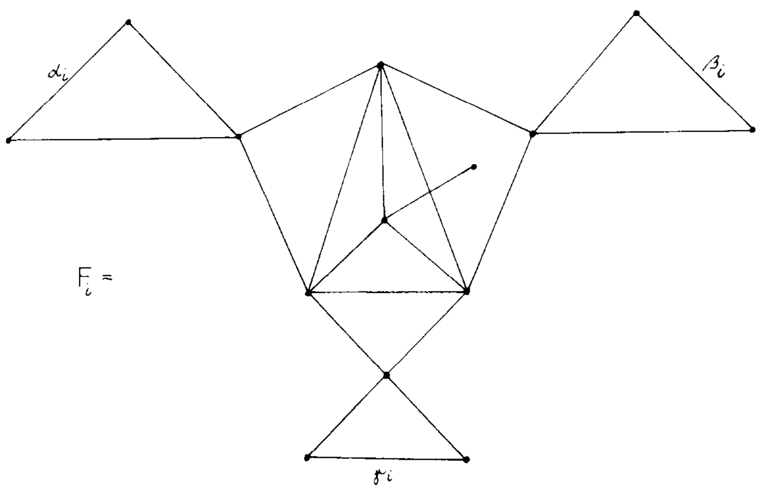

Fig. 5.

It is easy to see that $\tau(G) \leq \tau_d(G) \leq \tau(G)+1$. Nevertheless the following holds.

Theorem 2.5. *It is NP-hard to decide whether* $\tau(G)=\tau_d(G)$ *for a given graph G.*

Proof. We shall reduce to it the problem of determination of τ, which is NP-complete by the Theorem 2.1. Let a graph G be given. We shall consider two cases.

(1) let $\tau_d(G)=\tau(G)$. Consider graphs E_i, $i=1,\dots,|V(G)|$, given by Fig. 6. Evidently, $\tau(E_i)=i$, $\tau_d(E_i)=i+1$. Let $G_i=G+E_i$ and put

$$i_0=\max\{i \mid \tau_d(G_i)=\tau(G_i)\}.$$

Then obviously $\tau(G)=i_0$. (Since $\tau(G+E_i)=\max(\tau(G), \tau(E_i))$.)

(2) let $\tau_d(G)=\tau(G)+1$. Put $G_i=G+K_{1,i}$, $i=1,\dots,|V(G)|$, where $K_{1,i}$ is the i-star. Obviously $\tau(K_{1,i})=\tau_d(K_{1,i})=i$. Set $i_0=\max\{i \mid \tau_d(G_i)=\tau(G_i)\}$. Then $\tau(G)=i_0$.

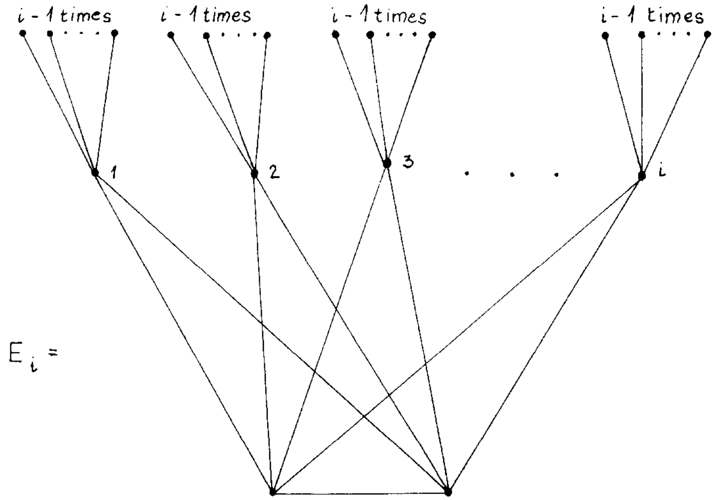

Fig. 6.

3. Set representations with a small universum set

For a given graph G denote by $\omega(G)$, $\omega_d(G)$ and $\omega_s(G)$ the minimum k for which there exists a set representation $\mathscr{F}$, a distinct set representation $\mathscr{F}$, and a simple set representation $\mathscr{F}$, respectively, such that $|\bigcup \mathscr{F}| = k$.

For a fixed integer k it is polynomial to decide whether $\omega(G) \leq k$, $\omega_d(G) \leq k$ and $\omega_s(G) \leq k$. For ω_d and ω_s it is enough to check all possible configurations. To prove it for ω define an equivalence relation $\sim$ on the vertices of G by

$$x \sim y \quad \text{iff} \quad (x,y) \in E \text{ and } ((x,z) \in E(G) \Leftrightarrow (z,y) \in E(G) \text{ for every } z \in V(G) - \{x,y\}).$$

Denote by $G/\sim$ the factorization of G by $\sim$. Let us note that $G/\sim$ is isomorphic to an induced subgraph of G. Since every set representation of $G/\sim$ is distinct, we get

$$\omega(G) = \omega(G/\sim) = \omega_d(G/\sim).$$

It is known that the problem of determination of ω (and hence also of ω_d) is NP-complete ([15]). We prove this for ω_s.

Theorem 3.1. *For a given graph G and an integer k it is NP-complete to decide whether $\omega_s(G) \leq k$.*

Proof. We shall reduce to it the problem of maximum independent set in a cubic graph [6]. First, let us state some observations about ω_s.

(1) Denote by $K_6 - e$, $K_6 - 2e$ and $K_6 - 3e$ the complete graph K_6 without one, two and three

disjoint edges, respectively. It is easy to check that

$$\omega_s(K_6-e)=\omega_s(K_6-2e)=5, \quad \text{and} \quad \omega_s(K_6-3e)=4.$$

(2) Let $G*H$ be a graph arising from G and H by glueing in a common edge $\bar{e}$. Then

$$\omega_s(G*H)=\min(\omega_s(G-\bar{e})+\omega_s(H),\ \omega_s(G)+\omega_s(H-\bar{e})).$$

Let $G=(V,E)$ be a cubic graph. For every vertex x of G consider a labeled graphs H_x given by Fig. 7, where e_x^1, e_x^2, e_x^3 are the edges of G incident to x.

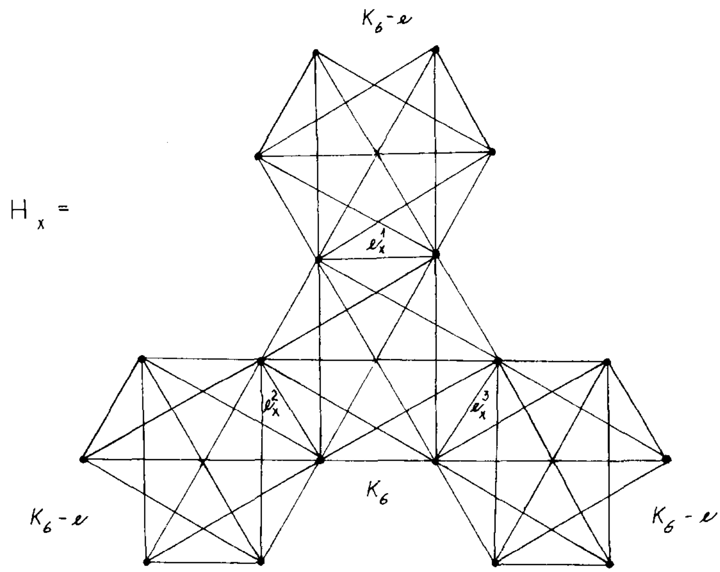

Fig. 7.

Using (1) and (2) we compute

$$\omega_s(H_x)=16,$$

$$\omega_s(H_x-e_x^1)=\omega_s(H-\{e_x^1,e_x^2\})=\omega_s(H-\{e_x^1,e_x^2,e_x^3\})=19.$$

Let H be an amalgamation of $H_x, x\in V(G)$, given by glueing the edges of the same label. Thus,

$$\omega_s(H)=16\cdot\alpha(G)+19(|V|-\alpha(G))=19|V|-3\cdot\alpha(G),$$

where $\alpha(G)$ denotes the maximum number of independent vertices of G.

4. Minimal forbidden induced subgraphs

The line graphs (i.e. the intersection graphs of graphs) are characterized by a finite family of minimal forbidden induced subgraphs. (See [1]). For the graphs which are intersection graphs of k-hypergraphs ($k>2$) the analogous statement does not hold. If we denote by Forb_k the class of all graphs G with $\tau(G)>k$ and $\tau(H)\le k$ for all subgraphs H induced on a proper subset of $V(G)$, then one can prove that $|\mathrm{Forb}_k|=\infty$. The same holds for the class $\mathrm{Forb}_k^{\mathrm{d}}$, which is defined in the same way but using distinct set representations only. An example of an infite class of graphs which belong to the Forb_3 and also to the $\mathrm{Forb}_3^{\mathrm{d}}$ is given by Fig. 8.

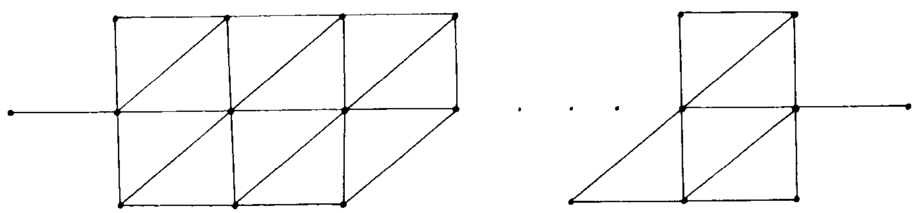

Fig. 8.

It would be interesting to know a nontrivial description of classes Forb_k for $k>2$. This may be difficult. The following problem seems to be easier: Is it true that there exists a constant c such that for all graphs of class Forb_3 the size of cliques is bounded by c?

For the class $\mathrm{Forb}_3^{\mathrm{d}}$ this question has a negative answer. An example of graphs from $\mathrm{Forb}_3^{\mathrm{d}}$ with aritrarily large cliques is given by Fig. 9.

5. Some remarks on random graphs

Let $0<p<1$ be fixed and denote by G_n a random graph with vertex set $\{1,2,\dots n\}$ such that each edge occurs with probability p independently of all other edges.

Proposition 5.1.

$$\mathrm{Prob}\,[\tau(G_n)=\tau_{\mathrm{d}}(G_n)]\to 1, \quad \text{and} \quad \mathrm{Prob}\,[\omega(G_n)=\omega_{\mathrm{d}}(G_n)]\to 1 \quad \text{as } n\to\infty.$$

For the proof it is enough to show that almost all graphs have the following property: The neighbourhoods $N_{x_1}, N_{x_2}, \dots, N_{x_n}$ of all vertices $x_1, x_2, \dots, x_n \in V(G_n)$ are pairwise distinct. Indeed,

$$\mathrm{Prob}\,[\exists x_1, x_2 \in V(G_n): N_{x_1}=N_{x_2}] \le \binom{n}{2}\cdot(p^2+(1-p)^2)^2)^{n-2}\to 0.$$

From the results of [7] where it is proved that the number of vertices of the largest complete subgraph of G_n is (with probability tending to 1)

$$\frac{2}{\log(1/p)}\log n + \mathrm{o}(\log n) \quad \text{as } n\to\infty,$$

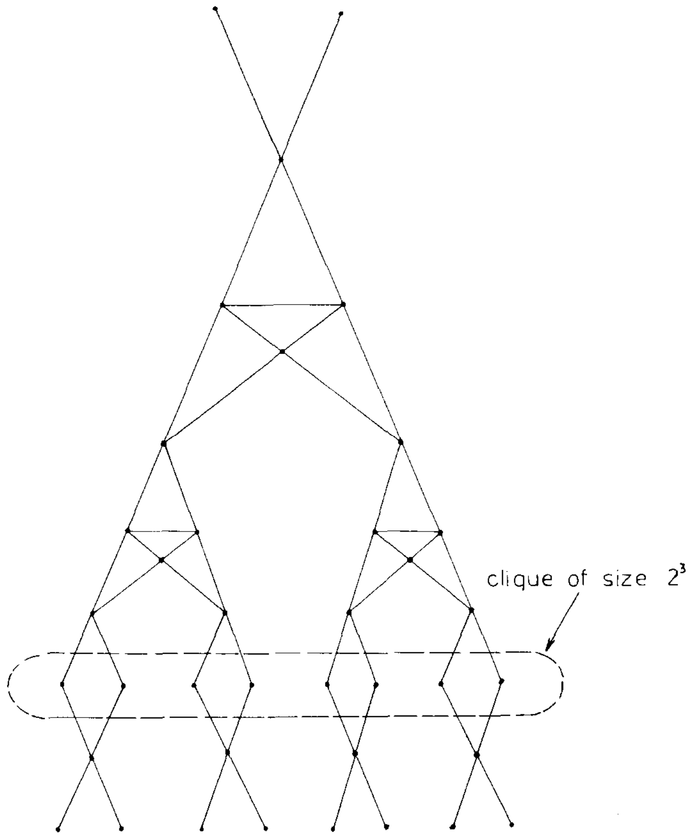

Fig. 9.

and the chromatic number $\chi(G_n)$ is (with probability tending to 1) at least

$$\frac{1}{2}\log\frac{1}{p}\cdot\frac{\mathrm{n}}{\log n}+\mathrm{o}\left(\frac{\mathrm{n}}{\log n}\right) \quad \text{as } n\to\infty,$$

it follows that the existence of constants c_1, c_2, d (depending on p only) such that

$$\text{Prob}\,[c_1 n^2/\log^2 n \le \omega(G_n) \le c_2 n^2/\log n] \to 1 \quad \text{as } n\to\infty,$$

$$\text{Prob}\,[d/\log n < \tau(G_n)/n < p] \to 1 \quad \text{as } n\to\infty.$$

From the fact that there exists an algorithm which colors G_n with at most $2\chi(G_n)$ colors with probability tending to 1 (see [7]), it follows the existence of an algorithm which covers the edges of G_n with probability tending to 1 by at most $cn/\log n$ cliques.

References

[1] L.W. Beineke, Derived graphs and digraphs, Beiträge zur Graphen-Theorie, Leipzig (1968) 17–33.
[2] C. Berge, Graphs and Hypergraphs (North-Holland, Amsterdam, 1973).
[3] J.C. Bermond and J.C. Meyer, Graphs représentent des arces d'un multigraphe, J. Math. Pures Appl. 9 52 (1973) 229–308.
[4] G. Erlich, S. Even, and R.E. Tarjan, Intersection graphs of curves in the plane, J. Combinatorial Theory (B) 21 (1976) 8–20.
[5] P. Erdös, A.W. Goodman, and L. Posa, The representation of graphs by set intersections, Canad. Math. J. 18 (1966) 106-612.
[6] M.R. Garey, D.S. Johnson, and L. Stockmayer, Some simplified NP-complete graph problems, Theoret. Comput. Sci. 1 (1976) 237-267.
[7] G.R. Grimmet and C.J.H. McDiarmid, On colouring random graphs, Math. Proc. Camb. Phil. Soc. (1975) 77-313.
[8] D.R. Fulkerson and D.A. Gross, Incidence matrices and interval graphs, Pacific J. Math. 15 (1965) 835–855.
[9] F. Harary, Graph Theory (Addison-Wesley, Reading, MA, 1969).
[10] M.R. Garey and D.S. Johnson, Computers and Intractability, (Freeman, San Francisco, 1979).
[11] R.M. Karp, Reducibility among combinatorial problems, in: R.E. Miller and J.W. Thatcher, eds., Complexity of Computer Computation (Plenum, New York, 1972) 85–103.
[12] E. Marzewski, Sur deux proprietes des classes d'ensembles, Fund. Math. 33 (1945) 303–307.
[13] C.H. Papadimitrou, The NP-completeness of the bandwidth minimization problem, Comput. 16 (1976) 263–270.
[14] S. Poljak and V. Rödl, Set systems determined by intersections, Discrete Math. 34 (1981) 173–184.
[15] L.T. Kou, L.J. Stockmeyer, and C.K. Wong, Covering edges by cliques with regard to keyword conflicts and intersection graphs, Comm. ACM (1978).