Can. J. Math., Vol. XXXIV, No. 3, 1982, pp. 519–524

A POLYNOMIAL ALGORITHM FOR CONSTRUCTING A LARGE BIPARTITE SUBGRAPH, WITH AN APPLICATION TO A SATISFIABILITY PROBLEM

SVATOPLUK POLJAK AND DANIEL TURZÍK

1. Introduction. Let G be a symmetric connected graph without loops. Denote by $b(G)$ the maximum number of edges in a bipartite subgraph of G. Determination of $b(G)$ is polynomial for planar graphs ([**6**], [**8**]); in general it is an NP-complete problem ([**5**]). Edwards in [**1**], [**2**] found some estimates of $b(G)$ which give, in particular,

$$b(G) \geqq f(m, n)$$

for a connected graph G of n vertices and m edges, where

$$f(m, n) = \tfrac{1}{2}m + \tfrac{1}{2}\{\tfrac{1}{2}(n - 1)\},$$

and $\{x\}$ denotes the smallest integer $\geqq x$.

We give an $O(V^3)$ algorithm which for a given graph constructs a bipartite subgraph B with at least $f(m, n)$ edges, yielding a short proof of Edwards' result.

Further, we consider similar methods for obtaining some estimates for a particular case of the satisfiability problem. Let Φ be a Boolean formula of variables $x_1, \ldots, x_n$. The formula Φ is called *satisfiable* if there exists a set of values of the variables for which Φ is true. The formula Φ is said to be in *conjunctive form* if

$$\Phi = (y_{1,1} \vee \ldots \vee y_{1,k_1}) \wedge (y_{2,1} \vee \ldots \vee y_{2,k_2}) \wedge \ldots \wedge (y_{q,1} \vee \ldots \vee y_{q,k_q})$$

where each $y_{i,j}$ equals x_t or $\neg x_t$ for some $t = 1, \ldots, n$. It is known that:

A. It is an NP-complete problem to decide whether an arbitrary Φ is satisfiable ([**3**]).

B. If Φ is given in conjunctive form with at most two variables per clause (i.e., $k_i \leqq 2$ for all $i = 1, \ldots q$) there is a polynomial algorithm for determining satisfiability of Φ ([**4**]).

C. If $\Phi = \varphi_1 \wedge \ldots \wedge \varphi_q$ is in conjunctive form with at most two variables in each clause, denote by $s(\Phi)$ the maximum number of φ's which can be satisfied. It is NP-complete to determine $s(\Phi)$ for all such Φ; moreover, it is NP-complete to decide whether $s(\Phi) \geqq (7/10)q$ ([**5**]).

Received June 29, 1979 and in revised form December 15, 1981.

We give (Theorem 2) a lower bound for $s(\Phi)$ in terms of n, m and p where m is the number of clauses of Φ with two distinct variables and p is the number of those with only one variable. In fact, we find in polynomial time a set of values for the variables such that at least $\frac{3}{4}m + \frac{1}{2}p$ (and hence $\frac{1}{2}q$) of the clauses are satisfied; this contrasts dramatically with the result [**5**] just quoted.

2. The main theorem. Let $G = (V, E)$ be a graph whose edge set E is partitioned into two disjoint sets E_1, E_2 called colors. Then $c = (E_1, E_2)$ is a 2-coloring of E, and (G, c) is an edge-2-colored graph. Any subset V_0 of the vertex set V determines a subgraph $H = H(G, c, V_0)$ whose vertex-set $V(H)$ is all of V and whose edge-set is

$$E(H) = \{(u, v) \in E_1 |\ u \in V_0, v \in V - V_0\} \cup \\ \cup \{(u, v) \in E_2 |\ u, v \in V_0 \quad \text{or} \quad u, v \in V - V_0\}.$$

H is called a g-bipartite or generalized bipartite subgraph of (G, c); a bipartite subgraph is the special case wherein $c = (E, \emptyset)$. We then define $h(G, c, V_0) = |E(H)|$, and

(1) $\quad b(G, c) = \max \{h(G, c, V_0) |\ V_0 \subset V\}.$

$b(G, c)$ is the maximum number of edges in a g-bipartite subgraph of (G, c), and of course

$$b(G, (E, \emptyset)) = b(G).$$

THEOREM 1. *If (G, c) is a simple connected edge-2-colored graph of n vertices and m edges, then*

$$b(G, c) \geqq f(m, n).$$

Proof. We prove the statement of the theorem by induction on n. For $n = 1$ the theorem trivially holds. Suppose $n > 1$, and let the theorem hold for all graphs with fewer than n vertices. We shall consider 3 cases.

Case 1. The graph G has an articulation vertex v.

Denote by $C_1, \ldots, C_k$ the connected components of $G - v$, by G_i the induced subgraph of G with vertex set $V(C_i) \cup \{v\}$, and by c_i the induced coloring of $E(G_i)$, $i = 1, \ldots, k$. Clearly,

$$b(G, c) = \sum_{i=1}^{k} b(G_i, c_i).$$

By the induction hypothesis $b(G_i, c_i) \geqq f(m_i, n_i)$, where $m_i = |E(G_i)|$ and $n_i = |V(G_i)|$, $i = 1, \ldots, k$. Since

$$\sum m_i = m \quad \text{and} \quad \sum (n_i - 1) = |V - v| = n - 1,$$

https://doi.org/10.4153/CJM-1982-036-8 Published online by Cambridge University Press

we obtain (using $\{x\} + \{y\} \geqq \{x + y\}$)

$$\sum_{i=1}^{k} f(m_i, n_i) \geqq f(m, n),$$

and hence $b(G, c) \geqq f(m, n)$.

Case 2. The graph G has no articulation vertex and there is a vertex v whose degree $d(v, G)$ in G is odd.

Then

$$b(G, c) \geqq b(G - v, c) + \tfrac{1}{2}(d(v, G) + 1).$$

By the induction hypothesis

$$b(G - v, c) \geqq f(m - d(v, G), n - 1).$$

Hence $b(G, c) \geqq f(m, n)$.

Case 3. The graph G has no articulation vertex and no vertex of odd degree.

Choose an arbitrary vertex u. There exists some edge (u, v) of G such that $G - \{u, v\}$ is connected. (To show this, we note that the blocks and articulation vertices of the graph $G - u$ form a tree (see [**7**], Theorem 4.4). If the tree is trivial (i.e., $G - u$ has no articulation vertex), then choose an arbitrary edge (u, v) incident with u; if not, consider any pendant block B of this tree, and the unique articulation vertex w of $G - u$ with $w \in V(B)$. Since w is not an articulation vertex of G, there must be a vertex $v \in V(B)$, $v \neq w$, such that (u, v) is an edge of G.)

Clearly,

$$b(G, c) \geqq b(G - u, c) + \tfrac{1}{2}d(u, G),$$

and as $d(v, G - u)$ is odd

$$b(G - u, c) \geqq b(G - \{u, v\}, c) + \tfrac{1}{2}(d(v, G - u) + 1).$$

Hence

$$b(G, c) \geqq b(G - \{u, v\}, c) + \tfrac{1}{2}(d(v, G) + d(u, G)).$$

By the induction hypothesis we obtain

$$b(G, c) \geqq f(m + 1 - d(v, G) - d(u, G), n - 2) + \tfrac{1}{2}(d(v, G) + d(u, G)) \geqq f(m, n).$$

COROLLARY 1 ([**2**]). *If G is a simple connected graph with m edges and n vertices then*

(2) $\quad b(G) \geqq f(m, n)$.

The proof follows from Theorem 1 by taking $c = (E, \emptyset)$.

https://doi.org/10.4153/CJM-1982-036-8 Published online by Cambridge University Press

COROLLARY 2. *If (G, c) is a simple edge-2-colored graph with m edges, n vertices, and k components, then*

$$b(G, c) \geqq f'(m, n, k)$$

where

$$f'(m, n, k) = \tfrac{1}{2}m + \tfrac{1}{2}\{\tfrac{1}{2}(n - k)\}.$$

Remark 1. The estimation (2) is the best possible e.g. for complete graphs and for amalgamation of complete graphs of odd order at one common vertex.

Remark 2. Since there is an algorithm for finding articulation vertices and blocks in $O(V^2)$ time (see [**9**], [**10**]), and since a suitable edge (u, v) in case 3 can be found in $O(V^2)$ time, the proof yields an $O(V^3)$ algorithm for finding a set $V_0 \subset V$ whose existence is shown by Theorem 1. In particular, for any connected simple edge-2-colored graph (G, c) there is a polynomial algorithm for finding a g-bapirtite subgraph of (G, c) with at least $f(m, n)$ edges.

Remark 3. $b(T_n, c) = n - 1$ for any tree T_n with n vertices and any edge-2-coloring c.

Remark 4. Let (G, c) be an edge-2-colored graph of m edges (not necessarily simple). Then there is a polynomial algorithm for constructing a set $V_0 \subset V(G)$ such that

$$(3) \quad h(G, V_0, c) \geqq \tfrac{1}{2}m,$$

and hence

$$b(G, c) \geqq \tfrac{1}{2}m.$$

Proof. Let $V(G) = \{v_1, \ldots, v_n\}$ and let G^i be the subgraph of G induced by the set $\{v_1, \ldots, v_i\}$, $i = 1, \ldots, n$. Set $V_0^1 = \{v_1\}$, and

$$V_0^i = \begin{cases} V_0^{i-1} \cup \{v_i\} & \text{if } h(G^i, V_0^{i-1} \cup \{v_i\}, c) \geqq h(G^i, V_0^{i-1}, c), \\ V_0^{i-1} & \text{otherwise.} \end{cases}$$

The set $V_0 = V_0^n$ satisfies (3).

Remark 5. Tarjan has previously made use of edge-2-colored graphs and generalized bipartite graphs in an investigation [**11**] of planar graphs.

3. Application to a satisfiability problem.

Let

$$\Phi = \varphi_1 \wedge \ldots \wedge \varphi_m \wedge \varphi_{m+1} \wedge \ldots \wedge \varphi_{m+p}$$

be a Boolean formula in conjunctive form of variables $x_1, \ldots, x_n$, where $\varphi_1, \ldots, \varphi_m$ are clauses with two distinct variables and $\varphi_{m+1}, \ldots, \varphi_{m+p}$ are clauses with only one variable. Assign to the formula Φ a graph G_Φ

https://doi.org/10.4153/CJM-1982-036-8 Published online by Cambridge University Press

with vertex set $\{x_1, \ldots, x_n\}$; let two distinct vertices x, y be joined by an edge $(x, y)_\varphi$ if there is a clause φ in Φ with variables x and y. Further define the edge-2-coloring $c_\Phi = (E_1, E_2)$ as follows:

$$(x, y)_\varphi \in E_1 \text{ if and only if } \varphi = (x \vee y) \quad \text{or} \quad \varphi = (\neg x \vee \neg y),$$
$$(x, y)_\varphi \in E_2 \text{ if and only if } \varphi = (\neg x \vee y) \quad \text{or} \quad \varphi = (x \vee \neg y).$$

THEOREM 2. *If Φ is a Boolean formula as above, and if G_Φ is a simple graph of k components, then*

$$(4) \quad s(\Phi) \geqq \tfrac{3}{4}m + \tfrac{1}{2}p + \tfrac{1}{4}\{\tfrac{1}{2}(n - k)\}$$

where $s(\Phi)$ is the maximum number of satisfiable clauses.

Proof. Let V_0 be a subset of the set $V = \{x_1, \ldots, x_n\}$ which realizes the maximum in the right hand side of (1). We can assign the value 0 to each variable of V_0 and the value 1 to each variable of $V - V_0$ or, conversely, the value 1 to each variable of V_0 and value 0 to each variable of $V - V_0$. In one of these cases at least

$$b(G_\Phi, c_\Phi) + \tfrac{1}{2}(m + p - b(G_\Phi, c_\Phi))$$

of the clauses must be satisfied. Thus

$$s(\Phi) \geqq \tfrac{1}{2}(m + p + b(G_\Phi, c_\Phi)).$$

By Corollary 2 of Theorem 1, $b(G_\Phi, c_\Phi) \geqq f'(m, n, k)$ and Theorem 2 follows.

COROLLARY 1. *Since $q = m + p$, we see that*

$$s(\Phi) \geqq \tfrac{1}{2}q + \tfrac{1}{4}\{\tfrac{1}{2}(n - k)\}.$$

Remark 5. If $x_1, \ldots, x_n, y_1, \ldots, y_n$ are variables and

$$\Phi = \bigwedge_{i \neq j} (x_i \vee x_j) \wedge \bigwedge_{i \neq j} (y_i \vee y_j) \wedge \bigwedge_{i,j} (\neg x_i \vee \neg y_j)$$

then the estimation stated by Theorem 2 is the best possible.

Remark 6. The proof of Theorem 2 gives a polynomial algorithm for finding values of the variables which realize the stated bound.

Remark 7. If G_Φ is not simple, we obtain

$$s(\Phi) \geqq \tfrac{3}{4}m + \tfrac{1}{2}p$$

using Remark 4 instead of Corollary 2 in the proof of Theorem 2.

Remark 8. Theorem 1 can be generalized to a "weighted" form:

Let $G = (V, E)$ be a simple connected graph, $c = (E_1, E_2)$ be an edge-2-coloring, and w: $E \to R^+$ be a non-negative function defined on edges of G. Then there exists a subset $V_0 \subset V$ such that

$$\sum \{w(e) | e \in E_1, |e \cap V_0| = 1\} + \sum \{w(e) | e \in E_2, |e \cap V_0| = 0 \text{ or } 2\}$$
$$\geqq \tfrac{1}{2} \sum \{w(e) | e \in E\} + \tfrac{1}{4}t(G, w),$$

where

$$t(G, w) = \min \left\{ \sum_{e \in T} w(e) \,|\, T \text{ is a spanning tree of } G \right\}.$$

4. Final remarks. If values of the variables $x_1, \ldots, x_n$ were assigned at random, we would find that the expectation is

$$E[s(\Phi)] = \tfrac{3}{4}m + \tfrac{1}{2}p;$$

thus the last term of (4) represents the benefit of judicious choice of the variables.

It would be interesting to know whether our Theorem 2 can be generalized to give some results when G_Φ is not simple. If the answer is "yes", it would be interesting to know the largest value of the constant c such that it is polynomial to determine whether $s(\Phi) \geqq c.q$.

The answers to these questions, like the results of the present paper, help to define the boundary between problems with polynomial solutions and NP-complete problems.

Acknowledgement. We are deeply grateful to the referee for many improvements of the paper.

References

1. C. S. Edwards, *Some extremal properties of bipartite subgraphs*, Can. J. Math. *25* (1973), 475–485.
2. ——— *An improved lower bound for the number of edges in a largest bipartite subgraph*, in: *Recent advances in graph theory* (Academia, Prague, 1975), 167–181.
3. S. A. Cook, *The complexity of theorem-proving procedures*, Proc. 3rd Ann. ACM Symp. on Theory of Computing, Association for Computing Machinery, New York (1970), 151–158.
4. M. Davis and H. Putnam, *A computing procedure for quantification theory*, JACM *7* (1960), 201–215.
5. M. R. Garey, D. S. Johnson and L. Stockmayer, *Some simplified* NP-*complete graph problems*, Theoretical Computer S ience *1* (1976), 237–267.
6. R. O. Hadlock, *Finding a maximum cut of a planar graph in polynomial time*, SIAM J. Comp. *4* (1975), 221–225.
7. F. Harary, *Graph theory* (Addison Wesley, Reading, Massachusetts, 1969).
8. G. I. Orlova and Y. G. Dorfman, *Finding the maximum cut in a graph*, Engrg. Cybernetics *10* (1972), 502–506.
9. K. Paton, *An algorithm for the blocks and cutnodes of a graph*, Comm. ACM *14* (1971), 468–476.
10. R. Tarjan, *Depth-first search and linear graph algorithms*, SIAM J. Comput. *1* (1972), 146–160.
11. ——— *An efficient planarity algorithm*, Thesis, Stanford University (1971).

Technical University,
Prague, Czechoslovakia;
Institute of Chemical Technology,
Prague, Czechoslovakia