

An Approach to Determination of Maximal Cliques in Undirected Graphs

S.V. Listrovoy

Doctor of Technical Sciences, Professor. S.V. Ukrainian State University of Railway Transport
(Ukraine, 61050, Kharkov, sq.Feyrbaha7, tel. (050) 9355042)
Email: om1sergeyvladimirovih@gmail.com

A.V. Sidorenko

Leading Software Engineer, Samsung Electronics Ukraine Company, LLC Samsung R & D Institute
(Ukraine, Kiev, 01302, str. Tolstoy 57, Tel. + 38050-98-00-852.)
Email: cdandrey@gmail.com

E.S. Listrovaya

Ph.D., Associate Professor of the National Aerospace University. N.E. Zhukovsky
(Ukraine, 61070, Kharkov, str. Chkalova 17.)
Email: listravkina@gmail.com

Received: 09 October 2017; Accepted: 29 November 2017; Published: 08 January 2018

Abstract—The article proposes the implicit exhaustive search procedure based on the triangle decomposition of graphs for determining the maximal clique in the arbitrary undirected graph G in polynomial time; it has allowed developing an exact algorithm for solving the problem with time complexity not exceeding $O(n^7)$, where n is the number of vertices in the graph G .

Index Terms—Maximal independent set, click, vertex cover, decomposition of a graph into triangles

I. INTRODUCTION

The Maximum Clique Problem (*MCP*) is one of the well-known NP-hard problems in the graph theory. There are not algorithms for solving it in polynomial time. Nevertheless, the problem has numerous applications. In bioinformatics *MCP* is applied in the computer analysis of genomic databases (e.g., search for potential regulatory structures of ribonucleic acids). In social networks *MCP* is used in database clustering to divide various communities into groups (clusters) of common properties. Clusterization allows processing each of them by a secondary server. In chemistry *MCP* is the basis of search for ‘the maximum common substructure’ in the graph describing the chemical compound structure. Moreover, *MCP* is a mathematical model for multiple problems linked with automation of electronic equipment engineering. Generally, the applications mentioned above, require exact solutions for *MCP*. And the input data volume is enormous (input graphs can contain up to a million of vertices). Thus, current research direction *MCP* is a development of new approaches towards finding exact solutions considering peculiarities of graphs in

applications. The maximum clique problem for undirected graphs is one of the combinatorial NP-complete problems. Generally, there is no polynomial algorithm for the problem [1]. And a great number of articles are dedicated to the problem [2]. All algorithms for solving *MCP* are classified into exact and approximate. The exact algorithms function, at the worst, during the time in exponential dependence on the input data volume. Clique-based algorithms (e.g., Cone [3], Burstall [4] and Raymond [5]) and backtracking algorithms (e.g., McGregor [6] and Wong [7]) are singled out among exact algorithms. Genetic algorithms [8] and combinatorial optimization algorithms [9] are singled out among the approximate ones.

The best known algorithms for finding the exact solution of *MCP* are the Bron-Kerbosh and the Wilf algorithms. The Bron-Kerbosh algorithm is a recursive procedure which upgrades a current clique considering one vertex at each step. The algorithm adds this vertex to the clique or includes it to the set of excluded vertices (which cannot be included in the clique). It is demonstrated, that the running time for the Bron-Kerbosh algorithm is $O(\text{poly}(n) \cdot 3^{n/3}) = O(\text{poly}(n) \cdot 1.4422^n)$, where $\text{poly}(n)$ is a polynomial of $n = |V|$. The Bron-Kerbosh algorithm has repeatedly been perfected. The fastest modification of the algorithm finds an exact solution of *MCP* in a time of $O(\text{poly}(n) \cdot 2^{0.249n}) = O(\text{poly}(n) \cdot 1.1888^n)$. The Wilf algorithm is also a recursive procedure. However, it finds not the maximum clique, but the largest independent set of vertices of the graph. It has been demonstrated that the running time for the Wilf algorithm is $O(\text{poly}(n) \cdot 1.39^n)$ $O(\text{poly}(n) \cdot 1.39n)$. Among the

approximate methods of solving the *MCP*, it is possible to single out papers [10-13]

II. BASIC CONCEPTS AND DEFINITIONS

Let $G = (V, E)$ be an undirected finite graph with the vertices V and the set of edges E , $|V| \geq 1$ and $|E| \geq 0$. The set of all vertices of the graph G , adjacent to the vertex $x \in V$, forms in G the environment of the vertex x specified by $N(x)$. The set of vertices $V' \subseteq V$ is called a clique of the graph G if in the graph $G(V')$ any two vertices are adjacent, and a maximal clique, if it is not included in a clique of larger number of vertices. The size of the largest clique (by number of vertices) of the graph G is specified by $\varphi(G)$ and called the density (or the clique number) of the graph G . The graph $G = (V, E)$, which is a clique, is called a complete graph and specified by K_n , where $n = |V|$. For it $\varphi(K_n) = |V|$. Generally, for the graph $G = (V, E)$ always $1 \leq \varphi(G) \leq |V|$. The subset $V' \subseteq V$ is assumed independent in the graph $G = (V, E)$ if the subgraph $G(V')$ does not contain edges. The largest independent set in terms of size is called maximal. The number of vertices in the maximal independent set is the independence number of the graph G designated by $\alpha_0(G)$. The independence number of the graph G and density of the complementary graph \bar{G} are connected by the formula: $\alpha_0(G) = \varphi(\bar{G})$, $\varphi(G) = \alpha_0(\bar{G})$. The identification variant of *MCP* can be formulated as follows.

STATEMENT. Given the graph $G = (V, E)$ and the positive number $K \leq |V|$.

QUESTION. Is it true that G contains a clique of the size exceeding K ? Otherwise, is there the subset $V' \subseteq V$, so that $|V'| \geq K$ and any two vertices in V' are connected with an edge from E ?

It is easy to demonstrate that the presence of an independent set of the size exceeding K in a complementary graph is the necessary and sufficient condition for existence in G a clique of the size K . Thus, knowing an exact solution about the problem of an independent set of vertices, it is possible to specify an exact solution to *MCP* and vice versa.

III. FORMALIZATION AND PROBLEM SOLUTION

Let us consider the triangle decomposition of a graph by an example of the complete 5-node graph G (Fig.1) and build a triangle on the base of each edge (i, j) , and the vertex number i is always lower than the vertex number j . The triangles built are presented in Figures 2-5.

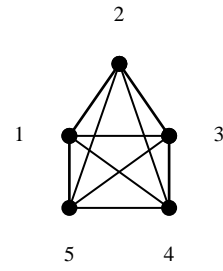


Fig.1. Graph G

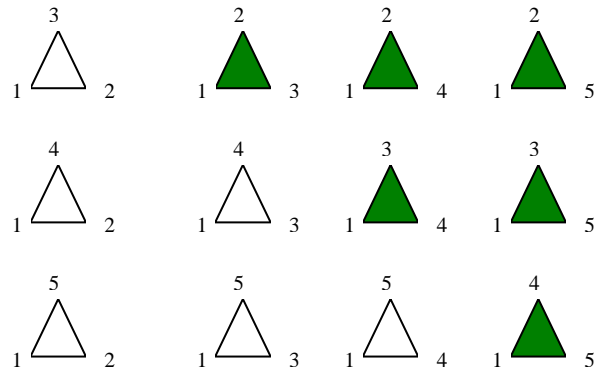


Fig.2. The set $\Omega_{i=1}$ of the triangles Δ_i built on the edges $(i = 1, j); j = 2,3,4,5$

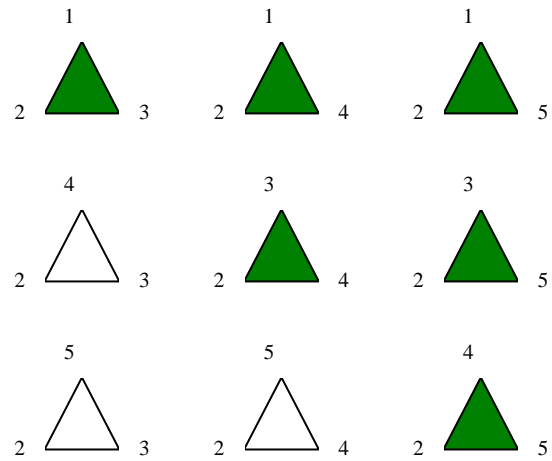


Fig.3. The set $\Omega_{i=2}$ of the triangles Δ_i built on the edges $(i = 2, j); j = 3,4,5$

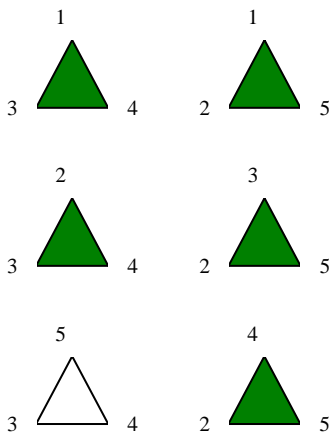


Fig.4.The set $\Omega_{i=3}$ of the triangles Δ_i built on the edges $(i = 3, j) ; j = 4,5$

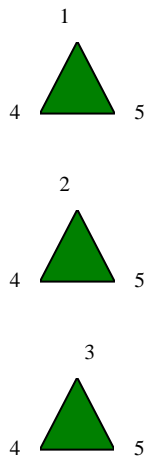


Fig.5.The set $\Omega_{i=4}$ of the triangles Δ_i built on the edge $(i = 4, j = 5)$

Subsequently, triangles will be specified by three figures indicating numbers of triangle vertices, for example, the first triangle $\Delta_{i=1}$ in Fig. 2 is specified as 123, where (12) is the edge on the base of which the triangle with the vertex 3 is formed. In Figures 2-5 the triangles which duplicate the triangles already build are highlighted and will be removed. Thereafter, the following triangles will remain and they are presented in Table 1.

Table 1.

Set $\Omega_{i=1}$ of the triangles	Set $\Omega_{i=2}$ of the triangles	Set $\Omega_{i=3}$ of the triangles
$\Delta_{\sigma=1}$	$\Delta_{\sigma=2}$	$\Delta_{\sigma=3}$
123 134 145	234 245	345
124 135	235	
125		

The sets Ω_i , as is seen from Table 1, are triangle subsets, in which we will distinguish the lines $S_j(\Omega_i)$ and columns $H_p(\Omega_i)$. So, for example, in the set $\Omega_{i=1}$ the triangles 123;134;145 form the first line $S_{j=1} = (\Omega_{i=1})$ of the triangle set $\Omega_{i=1}$, triangles 124;135 form the second line $S_{j=2} = (\Omega_{i=1})$ and the triangle 125 forms the third line $S_{j=3} = (\Omega_{i=1})$ of the triangle set $\Omega_{i=1}$. It is easy to note that if the size of a clique is n , the number of lines is $\lambda = n - 2$ in $\Omega_{i=1}$. The first column $H_{p=1}(\Omega_{i=1})$ in the set $\Omega_{i=1}$ is formed by the triangles 123;124;125, the second column $H_{p=2}(\Omega_{i=1})$ is formed by the triangles 134;135 and the third column $H_{p=3}(\Omega_{i=1})$ is formed by the triangle 145, their number φ is also equal to $n - 2$. When the number i increases by a unit of the set Ω_i the number of lines and columns also decreases by one. If the set of all triangles of the set Ω_i is recorded in accordance with their record in the columns Ω_i , we can say that the ordered set Ω_i is specified. Thus, the ordered set $\Omega_{i=1}$ (Table 1) is as follows

$$\Omega_{i=1} = \{123;124;125;134;135;145\}$$

Define the general number of various triangles for the case when a complete graph has n vertices. Specify the maximum number of triangles with the edge (1,2) by k , generally $k = n - 2$. Therefore, the general number of various triangles, which can be built in a complete graph is equal to the sum of the following lines.

$$z = (1 + 2 + \dots + k) + (1 + 2 + \dots + k - 1) + \dots + (1 + 2 + \dots + (k - 3) + \dots + 1) \tag{1}$$

Present the sum (1) as the following triangle for $n = 5$ and $k = n - 2 = 3$

$$\begin{aligned} &1 + 2 + 3 = 6 \\ &1 + 2 = 3 \\ &1 \\ &\text{the sum is } 10 \end{aligned}$$

It follows from the triangle that for the arbitrary k , if summarize the lines defined by relation (1) we form k ones, $k - 1$ twos, $k - 2$ trees, $k - 3$ fours, etc., that is the sum (1) can be presented as

$$z = 1k + 2(k - 1) + 3(k - 2) + 4(k - 3) + 5(k - 4) + \dots \quad (2)$$

Rewrite (2) as follows

$$\begin{aligned} z &= (k + 2k + 3k + 4 + 5k + \dots) - (2 \cdot 1 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 + \dots) = \\ &= k(1 + 2 + 3 + \dots + k) - [(1+1) \cdot 1 + (1+2) \cdot 2 + (1+3)3 + (1+4)4 + \dots = \\ &= k(1 + 2 + 3 + \dots + k) - [(1 + 2 + 3 + \dots + k - 1) + (1^2 + 2^2 + 3^2 + \dots + (k - 1)^2)] \end{aligned} \quad (3)$$

By considering

$$\begin{aligned} (1 + 2 + 3 + \dots + k) &= \frac{k}{2}(k + 1); \\ 1^2 + 2^2 + 3^2 + \dots + k^2 &= \frac{k}{6}(k + 1)(2k + 1) \end{aligned}$$

we will have

$$z = \frac{k}{2} \left[k(k + 1) - (k - 1) \left(1 + \frac{2k - 1}{3} \right) \right]$$

by substituting $k = n - 2$, finally we get

$$\begin{aligned} z &= \frac{(n - 2)}{2} \left\{ (n - 2)(n - 1) - (n - 3) \left[1 + \frac{2n - 5}{3} \right] \right\} = \\ &= \frac{n(n - 1)(n - 2)}{6}. \end{aligned} \quad (4)$$

Check it and suppose $n = 5$, so we will get the unknown quantity equal to 10.

$$z = \frac{5 \cdot 4 \cdot 3}{6} = 10$$

If a graph is incomplete, it decomposes into the triangles Δ_σ and edges, as there may appear edges on the base of which it is impossible to build triangles. For example, for the graph G presented in Fig.6 decomposition into triangles and edges is as follows: $\Delta_i = \{128; 145; 154; 237; 278; 281; 238; 372; 56; 67; 783\}$.

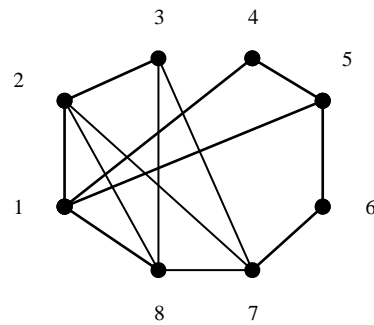
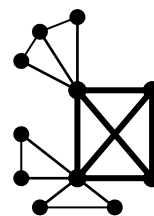


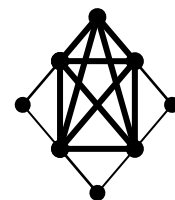
Fig.6. Graph G

Among the triangles the duplication triangles Δ_i are highlighted, and after their removing, we will obtain the set $\{\Delta_\sigma\} = \{128; 154; 237; 278; 238; 451; 56; 67; 783\}$ with $z = 7$ various triangles Δ_σ and $z^1 = 2$ edges. Some edges appearing in the process of decomposition of a graph, which cannot be the base to build triangles in the clique search, can be removed. It should be noted that the union of triangles Δ_σ in subsets Ω_i allows each to generate the largest clique Q_i . Choosing among them a clique of maximum dimension, we get the largest clique in the analyzed graph. This expansion will be used later to determine the maximum clique in arbitrary graphs.

The given characteristic of decomposition will further be used for determination of the maximal clique in arbitrary graphs. Decomposition of the graph, presented a clique, into triangles based on each edge (i, j) , in which the number of the vertex i will always be lower than number of the vertex j , creates the triangle subsets $\{\Omega_i\}$ which will be presented orderly hereafter. In an arbitrary graph when forming subsets among triangles, the union of which gives the maximal clique with the vertex i , there exist triangles not belonging to the given clique. They can be formed by joining the clique by separating vertices or built on the edges the clique (Fig.8). The vertex is called separating if its removal increases the connectivity component number in the graph.



Joining the triangles to the clique through the separating vertices



Joining the triangles through the clique edges

Fig.7. Alternatives of joining the triangles to the clique of 4 and 5 vertices

Thus, a set of triangles in the subsets Ω_i can be divided into two subsets $\{\Delta_\sigma^\gamma\}$ and $\{\Delta_\sigma^*\}$, where $\{\Delta_\sigma^\gamma\}$ is the set of triangles, the union of which forms in Ω_i the maximal clique Q_i^{\max} of the size γ ; $\{\Delta_\sigma^*\}$ is the subset of triangles linked to the clique Q_i^{\max} . With respect to the two subsets the following statements are true.

Statement 1. If in the subsets Ω_i there exists Q_i of the size $P(Q_i)$ on the set of vertices belonging to the triangles Δ_σ , so there exists the union of triangles $\Delta_\sigma \in \{\Delta_\sigma^\gamma\}$ which forms the given clique Q_i .

The trueness of this statement results from the triangle decomposition of a graph considered.

Statement 2. The union $\cup \Delta_\sigma$ of the arbitrary triangles $\Delta_\sigma \in \{\Delta_\sigma^\gamma\}$ forms a complete subgraph.

Let the union $\cup \Delta_\sigma$ of triangles form the clique Q_i of the size $P(Q_i)$. If we consider the union of two triangles $\Delta_{\sigma=1}$ and $\Delta_{\sigma=2}$, so that the union of the given triangles forms the complete subgraph, the set of all vertices of the first triangle should be linked with the set of vertices of the second triangle. It is obvious that the triangles forming the clique should satisfy the given characteristic. Suppose that there is the triangle $\Delta_\sigma^* \in \{\Delta_\sigma^\gamma\}$ which by joining the triangle Δ_σ^γ forms an incomplete graph, i.e., if we unite all triangles there will be at least one vertex not connected to all vertices, that is, the clique size is $P(Q_i) - 1$, which contradicts the assumption about the existence of the clique Q_i of the size $P(Q_i)$. The following consequence from the statement 2 is important.

Consequence. The union of the triangle $\Delta_\sigma^* \in \{\Delta_\sigma^\gamma\}$ and the union $\cup \Delta_\sigma$ of the arbitrary triangles $\Delta_\sigma \in \{\Delta_\sigma^\gamma\}$ lead to formation of an incomplete subgraph.

Therefore, if we unite two arbitrary triangles $\Delta_{\sigma=h}^\gamma \cup \Delta_{\sigma=p}^\gamma$ from the set $\{\Delta_\sigma^\gamma\}$, we will obtain a complete subgraph, and if we unite pairs of triangles $\Delta_{\sigma=k}^* \cup \Delta_{\sigma=d}^*$ from the subset $\{\Delta_\sigma^*\}$ or various subsets $\Delta_{\sigma=h}^\gamma \cup \Delta_{\sigma=d}^*$ we will obtain incomplete subgraphs. It is obvious, that an arbitrary union of triangles of the arbitrary subset of triangles $\Delta_\sigma \in \{\Delta_\sigma^\gamma\}$ always forms an incomplete graph. As the vertex enumeration in the graph under consideration is random, the maximal clique can be placed in one of the subsets Ω_i , a set of triangles

being in any of the subsets. In order to define the maximal clique in the graph under consideration, firstly, we need to find such unions of the triangles $\{\cup \Delta_\sigma\}$ which form the maximal cliques Q_i^{\max} in the subsets Ω_i , and, then, choose the maximum clique $Q_{\max}^i = \max\{Q_\sigma^{\max}\}$. Consider the formation of Q_i^{\max} from triangles of the subsets $\{\Delta_\sigma^\gamma\} \in \Omega_i$ and $\{\Delta_\sigma^*\} \in \Omega_i$. In terms of formation of cliques of triangles consider the following most important cases of unions of triangles in the subsets Ω_i for formation of Q_i^{\max} :

1. The triangles forming Q_i^{\max} in Ω_i do not cross the triangles forming other cliques in the graph G or there exist intersections only through separating vertices.
2. The triangles forming Q_i^{\max} in Ω_i intersect the triangles not belonging to any other clique of the graph G .
3. The triangles forming Q_i^{\max} in Ω_i intersect the triangles forming other cliques in the graph G of the size $P(Q_i) \leq P(Q_i^{\max})$.

Examples of the graphs corresponding to Cases (1-3) and their ordered sets Ω_i are presented in Table 2.

As is seen from Table 2, in Cases 1 and 2 there always exist the subsets Ω_i only with the triangles forming Q_i^{\max} , in Case 1 such subset is $\Omega_{i=1} = \{123;124\}$, and in Case 2 it is $\Omega_{i=1} = \{123;124;134\}$. In Case 3 the subset

$$\Omega_{i=1} = \{123;124;125;127;134;135;145;157\}$$

contains both the triangles $\{\Delta_\sigma^\gamma\} = \{123;124;125;135;145\}$ belonging to Q_i^{\max} and the triangles $\{\Delta_\sigma^*\} = \{127;157\}$ not belonging to the given clique. It raises the question about how Q_i^{\max} can be formed from subsets of the triangles $\{\Delta_\sigma^\gamma\} \in \Omega_i$ and $\{\Delta_\sigma^*\} \in \Omega_i$.

Therefore, take Statement 2 and the consequence from it, implying that if we unite two arbitrary triangles $\Delta_{\sigma=h}^\gamma \cup \Delta_{\sigma=p}^\gamma$ from the set $\{\Delta_\sigma^\gamma\}$, we will obtain a complete subgraph, and if we unite pairs of triangles $\Delta_{\sigma=k}^* \cup \Delta_{\sigma=d}^*$ from the subset $\{\Delta_\sigma^*\}$ or various subsets $\Delta_{\sigma=h}^\gamma \cup \Delta_{\sigma=d}^*$ we will obtain incomplete subgraphs. Start the building of Q_i^{\max} with enumeration of the triangles $\Delta_\sigma \in \Omega_i$; $i = (\overline{1, q})$ from 1 to t . The number t of the triangles Δ_σ in Ω_i cannot exceed the value $\eta = (n - 1)(n - 2)$. At $t=5$ the triangle sets T_σ is of the form (Table 3).

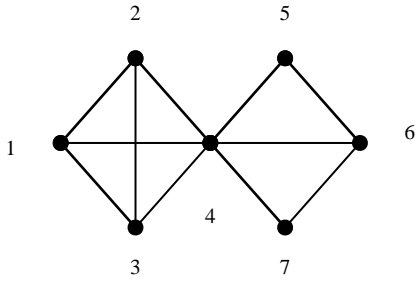


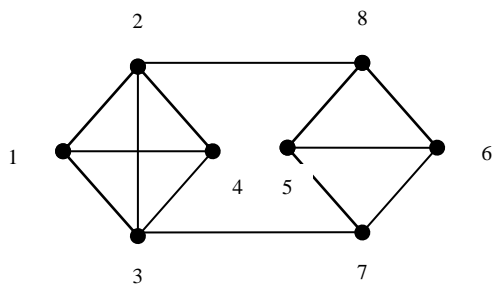
Table 2

Case 1

$$\Omega_{i=1} = \{123;124\} ; \Omega_{i=2} = \{234\}$$

$$\Omega_{i=4} = \{456;467\}$$

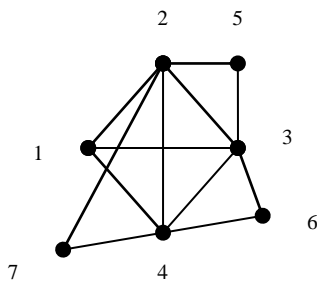
$$\Omega_{i=3} = \Omega_{i=5} = \Omega_{i=6} = \Omega_{i=7} = \emptyset$$



$$\Omega_{i=1} = \{123;124\} ; \Omega_{i=2} = \{234\}$$

$$\Omega_{i=5} = \{567;568\} ;$$

$$\Omega_{i=4} = \Omega_{i=6} = \Omega_{i=7} = \Omega_{i=8} = \emptyset$$

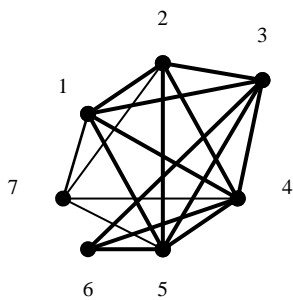


Case 2

$$\Omega_{i=1} = \{123;124;134\} ; \Omega_{i=2} = \{234;235;247\} ;$$

$$\Omega_{i=3} = \{346\}$$

$$\Omega_{i=4} = \Omega_{i=5} = \Omega_{i=6} = \Omega_{i=7} = \emptyset$$



Case 3

$$\Omega_{i=1} = \{123;124;125;127;134;135;145;157\}$$

$$\Omega_{i=2} = \{234;235;245;247;257\}$$

$$\Omega_{i=3} = \{345;346;356\}$$

$$\Omega_{i=4} = \{456;457\}$$

$$\Omega_{i=5} = \Omega_{i=6} = \Omega_{i=7} = \emptyset$$

Table 3

$T_{\sigma=1}$	$T_{\sigma=2}$	$T_{\sigma=3}$	$T_{\sigma=4}$	$T_{\sigma=5}$
1 2 3 4 5	2 3 4 5	3 4 5	4 5	5
1 3 4 5	2 4 5	3 5	5	
1 4 5	2 5	5		
1 5	5			
5				

The triangle set T_{σ} allows us to define all possible unions of triangles based on the triangle with the number $\sigma; \sigma = (\overline{1}, t)$ which give complete graphs in the union.

Each triangle of the set contains j lines, where $j = (\overline{1}, t)$. So, if the set of triangles $\{\Delta_{\sigma=1}; \Delta_{\sigma=2}; \dots; \Delta_{\sigma=t}\} \in \Omega_t$ contains t triangles, then the first line of the triangle set $T_{\sigma=1}$ also contains t triangles. Consider unification of triangles by example of this line; the unification of triangles in other lines runs similarly.

Procedure B of triangle unification of j-line in the triangle set T

Choose the triangle $\Delta_{\sigma=1} \in \Omega_i$ in j -line out of the triangle set $\{\Delta_{\sigma=1}; \Delta_{\sigma=2}; \dots; \Delta_{\sigma=t}\}$ and form the union $\Delta_{\sigma=1} \cup \Delta_{\sigma=2}$ which creates a subgraph (complete or incomplete). If the graph is incomplete the union $\Delta_{\sigma=1} \cup \Delta_{\sigma=2}$ is marked with (-) and the triangle $\Delta_{\sigma=2}$ is skipped, the union with the next triangle $\Delta_{\sigma=3}$ is checked and union $\Delta_{\sigma=1} \cup \Delta_{\sigma=3}$ (+) is marked with (+). Consider the union $\Delta_{\sigma=1} \cup \Delta_{\sigma=3} \cup \Delta_{\sigma=4}$ (+) and it gives the complete subgraph, then consider the union $\Delta_{\sigma=1} \cup \Delta_{\sigma=3} \cup \Delta_{\sigma=4} \cup \Delta_{\sigma=5}$ (+), that is, all triangles of the union belong to the subset $\{\Delta_i^y\}$ and it also gives the complete subgraph, then consider the union $\Delta_{\sigma=1} \cup \Delta_{\sigma=3} \cup \Delta_{\sigma=4} \cup \Delta_{\sigma=5} \cup \Delta_{\sigma=6}$ (-), i.e. the triangle $\Delta_{\sigma=6} \in \{\Delta_\sigma^\lambda\}$, skip it and begin forming the next union $\Delta_{\sigma=1} \cup \Delta_{\sigma=3} \cup \Delta_{\sigma=4} \cup \Delta_{\sigma=5} \cup \Delta_{\sigma=7}$ and repeat the procedure. After formation of all possible unions with the 1st triangle in j -line, select the union with the maximal clique Q_j^{\max} , defined by unification of all triangles belonging to $\{\Delta_\sigma^\lambda\}$. If $\Delta_{\sigma=1} \in \{\Delta_\sigma^*\}$ and all its unions give incomplete graphs, Q_j^{\max} is the empty set $\{\emptyset\}$.

By applying procedure **B** to all lines of the triangle sets $\{T_\sigma\}$, form the sets $\{Q_j^{\max}\}$ based on the lines of triangle sets and select among them the maximal clique $Q_{\max}^\sigma = \max\{Q_{j=1}^{\max}; Q_{j=2}^{\max}; \dots; Q_{j=t}^{\max}\}$ obtained on the base of j -line of the triangle set T_σ . Having formed the set of cliques $\{Q_{\max}^\sigma\}$ for all triangle sets $\{T_\sigma\}$ and selecting among them the maximal clique $Q_i^{\max} = \max\{Q_{\max}^1; Q_{\max}^2; \dots; Q_{\max}^t\}$ we obtain the maximum clique in the subset Ω_i . Further, we can similarly determine the cliques Q_i^{\max} for the other subsets $\Omega_i; i = (\overline{1, q})$ and among the subsets of cliques $\{Q_i^{\max}\}; i = (\overline{1, q})$ select the maximal clique $Q_{\max} = \max\{Q_i^{\max}\}$ and it turns out to be the maximum clique in the graph G . Thus, we can propose the following procedure **A** for determination of the maximal clique in an arbitrary graph.

Procedure A

Step 1. Form the ordered subsets of triangles Ω_i based on the edges of a graph beginning with the vertex of a lower number on the edges (i, j) on condition that $i > j$ and pass on to the next step.

Step 2. Form the triangle sets $\{T_\sigma\}$ in the ordered subsets Ω_i and, by applying procedure **B** to all lines of the triangle sets $\{T_\sigma\}$ of the subsets Ω_i , form the sets $\{Q_j^{\max}\}$ on the lines of the triangle sets and select among them the maximal cliques $Q_{\max}^\sigma = \max\{Q_{j=1}^{\max}; Q_{j=2}^{\max}; \dots; Q_{j=t}^{\max}\}$ and pass on to the next step.

Step 3. Among the subset of the formed cliques $\{Q_i^{\max}\}$ choose the maximal clique $Q_{\max} = \max\{Q_i^{\max}\}$ and the procedure finishes.

Consider an example of the procedure for the graph G presented in Fig.8.

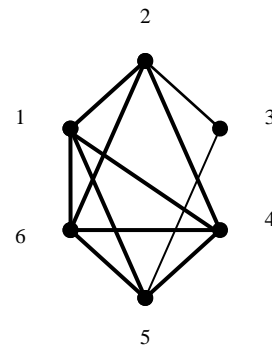


Fig.8. Graph G

Form the subsets Ω_i for the given graph, their form is

$$\Omega_{i=1} = \{124; 126; 145; 146; 156\}$$

$$\Omega_{i=2} = \{246\}; \Omega_{i=3} = \emptyset;$$

$$\Omega_{i=4} = \{456\}; \Omega_{i=5} = \emptyset; \Omega_{i=6} = \emptyset;$$

Renumber the triangles in the subset

$$\Omega_{i=1} = \{124; 126; 145; 146; 156\} \text{ (Table 4).}$$

Table 4

Triangle number	1	2	3	4	5
$\{\Delta_\sigma\} \in \Omega_{i=1}$	124	126	145	146	156

Form the triangle sets $\{T_\sigma\}$ and apply to them procedure **B** to unite the triangles (Table 5)

Table 5

$\{T_\sigma\}$	Procedure B
T_1	
1 2 3 4 5	112 +;123 -;124+;1245-; obtained the maximal union of triangles 124; the union of these triangles formed the complete subgraph on the vertices of united triangles, i.e. obtained the clique $Q_{j=1}^{\max} = \{1246\}$
1 3 4 5	13-;14+;145-; obtained the maximal union of triangles 14 and obtained the clique $Q_{j=2}^{\max} = \{1246\}$
1 4 5	14+;145-;obtained the maximal triangle union 14 and obtained the clique $Q_{j=3}^{\max} = \{1246\}$
1 5	15-; obtained $Q_{j=4}^{\max} = \emptyset$
5	$Q_{j=5}^{\max} = \{156\}$
T_2	
2 3 4 5	23-;24+;245-; obtained the maximal union of triangles 24 and obtained the clique $Q_{j=1}^{\max} = \{1246\}$
2 4 5	24+;245-;obtained the maximal triangle union 24; obtain the clique $Q_{j=2}^{\max} = \{1246\}$
2 5	25-; obtained $Q_{\sigma=3}^{\max} = \emptyset$
5	$Q_{j=4}^{\max} = \{156\}$
T_3	
3 4 5	34+;345; obtained the maximal triangle union 345; obtained the clique $Q_{j=1}^{\max} = \{1456\}$
3 5	35+; obtained the maximal triangle union 35; obtained the clique $Q_{j=2}^{\max} = \{1456\}$
5	$Q_{j=4}^{\max} = \{156\}$
T_4	
4 5	45+; obtained the maximal triangle union 45; obtained the clique $Q_{j=1}^{\max} = \{1456\}$
5	$Q_{j=2}^{\max} = \{156\}$
T_5	
5	$Q_{j=1}^{\max} = \{156\}$

For instance, the triangle union $1 \cup 2 = 124 \cup 126$ gives a complete subgraph (Fig.9a), and the triangle union

$1 \cup 2 \cup 3 = 124 \cup 126 \cup 145$ gives an incomplete graph (Fig.9b).

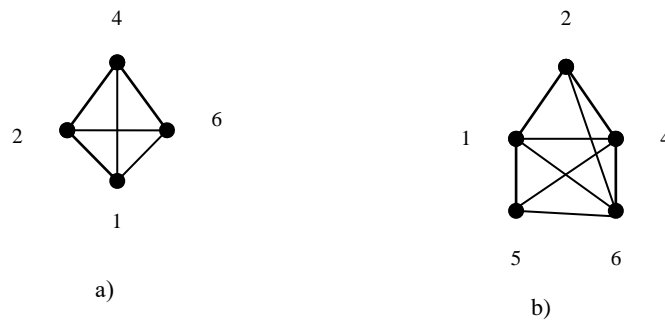


Fig.9. The subgraphs obtained by uniting the triangles 1,2,3.

Then, determine $Q_{i=1}^{\max} = \max\{1246;1456;156\}=1246$ or 1456; $Q_{i=2}^{\max} = \{246\}$; $Q_{i=4}^{\max} = \{456\}$, hence $Q_{\max} = \max\{1246;1456;356;456\} = 1246$ or 1456, i.e. the maximal clique is found and in this case there are two of them 1246 and 1456

IV. ASSESSMENT OF COMPLEXITY AND EFFICIENCY OF PROCEDURE A

Formation of the triangle subsets Ω_i is on the edges (i, j) incident to the vertex i , the number of edges incident to the vertex i cannot exceed $n-1$, and the number of triangles to be built on one edge cannot exceed $n-2$, therefore, the general number of triangles in the subset Ω_i do not exceed $h = (n-1)(n-2) < n^2$, the number of the sets Ω_i is $q \leq n-1$. Consequently, the general number of triangles to be built in procedure A does not exceed $(n-1)^2(n-2) \leq n^3$, taking into account that forming one triangle consists of less than $2n$ comparison operations, procedure A takes less than $2n^4$ comparison operations to form the subset Ω_i . Thereupon, procedure B begins to run in the subsets Ω_i and analyses the triangle sets $T_{\sigma=1}; T_{\sigma=2}; \dots; T_{\sigma=h}$, determines the number of unions for analysis to form the maximal clique. The most labour-intensive is the analysis of the first triangle set with the number of lines t equal to h . And the first line $T_{\sigma=1}$ also contains h triangles. The number of lines and operations of unions in a line is given in Table 6.

Table 6.

Triangle sets	Number of unions
$T_{\sigma=1}$	$\frac{h(h-1)}{2}$
$T_{\sigma=2}$	$\frac{(h-2)(h-3)}{2}$
...	...
$T_{\sigma=h-1}$	2
$T_{\sigma=h}$	1

Table 6 demonstrates that the number of unions is determined by the sum of the following sequence $1+2+\dots+h-1 = \frac{h(h-1)}{2}$. Similarly, the other triangle sets are presented in Table 7 which shows that the number of unification operations to analyze all triangle sets cannot exceed $0.5h^2(h-1) < 0.5h^3$.

Table 7

Triangle numbers				Number of unions	
1	2	3	...	h	$h-1$
	1	3	...	h	$h-2$
	
			$h-1$	h	2
				h	1

Less than $2n$ comparison operations will be needed so that to establish the fact whether the subgraph complete or incomplete while uniting the triangles, and h does not exceed n^2 . Consequently, considering the fact that we have to select the maximal element in the massif of the size not exceeding h , the general number of comparison operations in procedure B with triangle sets is equal to $0.5h^3 \cdot 2n + h \log h = n^6 + 2n^2 \log n \approx n^6$. Thus, for a maximal clique in one triangle set Ω_i less than n^6 comparison operations are required. As the number of subsets Ω_i cannot exceed $n-1$ and we need to select the maximal element out of the massif of numbers not exceeding $n-1$, the general number of operations in procedure A does not exceed $n^6 \cdot (n-1) + (n-1) \log(n-1) \approx n^7$. As procedure B performs an implicit complete search for triangle unions which can lead to complete subgraphs, procedure A always gives the exact solution.

V. EXPERIMENTAL RESEARCH OF PROCEDURE A

During the experimental research more than 50 problems were generated at each point. The confidence probability of the findings obtained was 0.95, the edge density in the graph changed from 0.3 to 0.6, the edge density of the graph was evaluated by the relation $\rho = \frac{m}{E_{\max}}$, where m is the number of edges in the complete graph $G(V,E)$, and $E_{\max} = \frac{n(n-1)}{2}$ is the maximum possible number of edges in the graph $G(V,E)$ with n -vertices. The numbers of vertices in the graph changed from 10 to 150, the edges were generated by the distribution uniform law. The range of parameters ρ and n is conditioned by the possibility to check the accuracy of the solutions obtained, since when ρ works for one, the procedure complexity works for the upper mark $O(n^7)$ and if n is more than 20 and the density is $\rho = 0.89$, it is almost impossible to check the accuracy of solution. Validation of the developed problem-solving procedure when analyzing the graph G was based on building the complementary graph \bar{G} . Then all maximal

independent sets for the graph \overline{G} with the algorithm from [14] were enumerated. Furthermore, the following was checked: whether the size of the maximal clique obtained in procedure A equal to the size of the greatest maximal set in the graph \overline{G} . If it is coincident, we can conclude that the problem is true. The verification did not reveal any inexact solutions. The findings of the experimental research

are demonstrated in Tables 8-11. In the Tables the parameter α is the average general number of various triangles built in procedure A; λ is the average size of the maximal clique built in procedure A; EO is the average number of elementary comparison operations in procedure A; t is the average time of procedure A in seconds; m is the average number of edges in the graph.

Table 8

Dependencies of average number of elementary comparison operations (EO), time of triangle formation (t) and their number α , as well as average value of time complexity in building various triangles on the density of edges ρ in the graph

$n = 40$										
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
EO	700	2423	5140	8340	12933	17498	22973	28675	35442	42640
t (sec)	0.013	0.1	0.3	0.85	1.56	2.65	4.29	6.49	9.73	14.54
α	7	67	232	655	1232	2116	3382	5040	7219	9880
$O(n^x)$	$n^{1.8}$	$n^{2.2}$	$n^{2.6}$	$n^{2.6}$	$n^{2.6}$	$n^{2.7}$	$n^{2.8}$	$n^{2.8}$	$n^{2.9}$	$n^{2.9}$

Table 9

Dependency of the average value: operations executed in procedure A (EO); time of procedure A; number of triangles built α ; dimensions of maximal cliques λ ; number of edges of the graph

$\rho = 0.3$

n	30	50	80	100	150
EO	5463	74961	1474957	9401679	114473122
t (sec)	0.09	0.2	1.58	8.81	103.48
α	96	489	2202	4316	14858
λ	4	5	6	6	6
m	130	367	948	1485	6352

Table 10

Dependency of the average value: operations executed in procedure A (EO); time of procedure A; number of triangles built α ; dimensions of maximal cliques λ ; number of edges of the graph

$\rho = 0.5$

n	10	20	30	40	45
EO	345	49928	1229337	95765414	781267305
t (sec)	0.02	0.14	1.97	136.5	911.6
α	13	142	487	1203	1754
λ	4	5	7	8	8
m	22	95	217	390	495

Table 11

Dependency of the average value: operations executed in procedure A (EO); time of procedure A; number of triangles built α ; dimensions of maximal cliques λ ; number of edges of the graph m ;

$\rho = 0.6$

n	15	20	25	30
EO	49928	1229337	95765414	781267305
t (sec)	0.14	1.97	136.5	911.6
α	142	487	1203	1754
λ	5	7	8	8
m	95	217	390	495

If the densities are in a range of 0.8-0.9, the complexity of procedure **A** is close to the upper value $O(n^7)$. As the analyses of the subsets Ω_i and triangle sets $\{T_\sigma\}$ can be fulfilled independently, procedure **A** can be efficiently parallelized which allows solving problems for graphs of great densities in real time.

VI. INFERENCE

Thus, for the first time a polynomial algorithm of the problem of maximal clique solution was built. Although the polynomial degree is rather high, the fact that a polynomial algorithm exists is more important for solving the task. As far as the task is one of the class NP-complete problems, the result obtained needs considerations. The existence of an algorithm of polynomial complexity considered in the study for determination of the maximal clique does not equalize the classes P and NP, whereas, as study [15] shows, Cook's theorem on the concept of NP-completeness of SAT is false. Study [15] puts forward the hypothesis that the existence at least one NP-complete problem is algorithmically unsolvable and the whole set of problems of discrete optimization and theory of graphs can be divided into subsets, in which one problem can be reduced to another polynomially. In such problem subsets the solution of one problem with the help of a polynomial algorithm leads to existence of the polynomial solution algorithm for the whole subset. But the question of polynomial transformation of problems for one such a subset into problems of other subsets is not solved. In this context such a subset forms problems as SAT and also the following: determination of graph isomorphism, determination of isomorphic embedding of a graph, determination of maximal independent sets and minimal vertex cover of graphs, as well as schedule feasibility, as they can easily be reduced to MCP in polynomial time.

REFERENCES

- [1] Kann, V., On the Approximability of NP-Complete Optimization Problems, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Sweden, 1992.
- [2] John W. Raymond and Peter Willett. Maximum Common Subgraph Isomorphism Algorithms Matching Chemical Structures *Journal of Computer-Aided Molecular Design*, 16: 521–533, 2002
- [3] Cone, M., Venkataraghavan, R. and McLafferty, F., *J. Am. Chem. Soc.*, 99 (1977) 7668.
- [4] Barrow, H. and Burstall, R., *Inf. Proc. Lett.*, 4 (1976) 83.
- [5] Raymond, J., Gardiner, E. and Willett, P., *Comput. J.*, in the press.
- [6] McGregor, J., *Software Pract. Exper.*, 12 (1982) 23.
- [7] Wong, A. and Akinniyi, F., *Proc. Int. Conf. Systems, Man and Cybern.*, Bombay & New Delhi, India, 1983, pp. 197.
- [8] Brown, R.D., Jones, G., Willett, P. and Glen, R., *J. Chem. Inf. Comput. Sci.*, 34 (1994) 63.

- [9] Funabiki, N. and Kitamichi, J., *IEICE Trans. Inf. & Syst.*, E82-D (1999) 1145.
- [10] Gribkov M., Alexeevski A., Ivanova D., Karyagina A., Spirin S. Life Core, program for classification of 3D structures of macromolecules // *Biophysics (Moscow)*. 2004. Vol. 48. Suppl. 1. P. 157–166.
- [11] Detecting highly overlapping community structure by greedy clique expansion / C. Lee, F. Reid, A. McDaid, N. Hurley // *Arxiv preprint arXiv:1002.1827*. _ 2010.
- [12] Uncovering the overlapping community structure of complex networks in nature and society / G. Palla, I. Derenyi, I. Farkas, T. Vicsek // *Nature*. 2005. _ V. 435, . 7043. _ P. 814–818.
- [13] Litvinenko V.A. Adaptive algorithms of definition of extreme sets of graphs // *Proceeding of the International Scientific Conferences «Intelligent System (IEEE AIS'03)» and «Intelligent CAD's (CAD-2003)»*. Scientific publication in 3 volumes. – 2003. – Vol. 3. – C. 52.
- [14] Listrovoy S.V. The method of enumeration of maximal independent sets in arbitrary non-oriented graphs // *Electron. Modeling-2014-36-No.1-C.3-17*.
- [15] Listrovoy S.V. On Correlation of P And NP Classes // *I.J.Modern Education and Computer Science*, 2012, 3, 21-27.

Authors' Profiles



Listrovoy Sergey Vladimirovich, Doctor of technical sciences, professor of Ukrainian State Academy of Railway Transport, Kharkov. In 1972 has finished high military command engineering school in Kharkov. The Area of the scientific studies of the problem to discrete optimization and graph theory and their use to analysis of the computing systems and networks.



Sidorenko Andrei Vladimirovich, Leading Software Engineer of Samsung Electronics Ukraine Company, LLC Samsung R & D Institute Ukraine. In 2001, he graduated from the Kharkov Military University with a degree in Radio Electronic Systems, Systems and Means of Armament of Military Equipment. The field of scientific research of the problem of discrete optimization and graph theory and their applications to the analysis of computer systems and networks.



Listrovaya Elena Sergeevna, Candidate of Technical Sciences, Associate Professor Department of Economics and Marketing National Aerospace University. N.E. Zhukovsky. Graduated in 1998, the National Aerospace University. N.E. Zhukovsky. Area of scientific interests application of information systems in the economic sphere of activity.

How to cite this paper: S.V. Listrovoy, A.V. Sidorenko, E.S. Listrovaya, "An Approach to Determination of Maximal Cliques in Undirected Graphs", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.10, No.1, pp. 1-12, 2018.DOI: 10.5815/ijmeecs.2018.01.01