

ANALYSIS OF HUBS LOADS IN BIOLOGICAL NETWORKS

*G.Sh. Tsitsiashvili, **A.S. Losev, ***M.A. Osipova, ****Yu.N. Kharchenko

IAM, FEB RAS, Vladivostok, Russia,
FAR Eastern Federal University

e-mails: *guram@iam.dvo.ru, **alexax@bk.ru,
mao1975@list.ru, *har@iam.dvo.ru.

ABSTRACT

Numerical experiments with biological networks are made and hubs loads are analysed. These experiments are based on algorithms of oriented graphs factorization. Numerical results allow to find narrow places in the networks.

INTRODUCTION

In articles [1], [2] we consider oriented graph with finite sets of vertices and edges. A sequential algorithm of graph factorization with square number of arithmetical operations by a number of graph vertices is used. A decrease of calculation complexity in a comparison with Boolean multiplication of contiguity matrixes is connected with an introduction of partial order matrix which is defined completely by assignment operations.

But a problem is to make numerical experiments, to analyze possibilities of suggested algorithm and to consider interesting protein networks with large numbers of nodes. For this aim we take protein network of Arabidopsis and analyze not only calculation complexity of the algorithm but find some new properties of considered network.

1. PRELIMINARIES

Consider oriented graph G with finite sets of vertices V and edges E . Say that two vertices $v_1, v_2 \in V$ of oriented graph belong to binary relation $v_1 \sim v_2$ if in this graph there is a cycle which contains both vertices. Call the oriented graph $[G]$ with the set of vertices $[V]$ and the set of edges $[E] = \{([v_1], [v_2]) : \exists v'_1 \in [v_1], v'_2 \in [v_2], (v'_1, v'_2) \in E\}$ the factor graph. It is obvious that $[G]$ is acyclic graph.

On the set $[V]$ of the graph $[G]$ vertices define the binary relation " \succeq ": $[v_1] \succeq [v_2]$ if in the graph $[G]$ there is a way from the cluster $[v_1]$ to the cluster $[v_2]$. It is obvious that if $[v_1] \succeq [v_2]$ then in the graph $[G]$ there is a way from any vertex of the cluster $[v_1]$ to any vertex of the cluster $[v_2]$.

Describe now an algorithm of a factorization of the graph $[G]$ vertices and a construction of partial order " \succeq ". Assume that the set $V = \{1, \dots, n\}$ consists of n vertices and denote $t = \min(f : 2^f \geq n) = \lceil \log_2 n \rceil + 1$ and put $A = \|a_{ij}\|_{i,j=1}^n$ the contiguity matrix of the graph G . Construct a sequence of zero-one matrixes $A(k) = \|a_{jr}(k)\|_{j,r=1}^n$ by the recurrent relation

$$a_{jr}(1) = \max_{1 \leq s \leq n} \min(a_{js}, a_{sr}), \quad a_{jr}(k+1) = \max_{1 \leq s \leq n} \min(a_{js}(k), a_{sr}(k)), \quad 1 \leq k < t.$$

It is obvious that if $a_{j'r}(t) = 1$ then in the graph G there is a way from the vertex j to the vertex r and $j' \in [j] \Leftrightarrow a_{j'r}(t) = a_{jj'}(t) = 1$. The relation $[i] \succeq [j] \Leftrightarrow \exists i' \in [i], j' \in [j]: a_{i'j'}(t) = 1$.

Cubic complexity of this algorithm is strong restriction for numerical experiments with oriented graphs occurred in an analysis of protein networks. So it is naturally to decrease calculation complexity.

Describe sequential algorithm of a construction of clusters and the matrix of partial order between clusters. On the step 1 there is the single vertex 1 which creates the cluster $[1]$ and the set of clusters $K = \{[1]\}$. Introduce the matrix $a = \|a([p], [q])\|_{[p], [q] \in K}$ which characterizes the partial order " \succeq ": $a([p], [q]) = 1$, if $[p] \succeq [q]$ and in opposite case $a([p], [q]) = 0$.

On the step 1 put $a([1], [1]) = 1$. Assume that on the step t there is the clusters set K and the matrix a . These clusters create a division of the set $V = \{1, \dots, n\}$ into non intersected subsets, each cluster $[k] \in K$ is indexed by maximal number $k \in V$ of its vertices.

On the step $t+1$ new vertex $t+1$ appears. It is connected with edges which come into vertices from the set $P \subseteq V$ and with vertices which come into the vertex $t+1$ from the vertices from the set $Q \subseteq V$. Each vertex from the set P (the set Q) comes into some cluster. Denote P the sets of clusters corresponding to the sets of vertices from P and define

$$K_{[p]} = \{[k] \in K : a([p], [k]) = 1\}, [p] \in P,$$

$$K_{[q]} = \{[k] \in K : a([k], [q]) = 1\}, [q] \in Q, A = \left(\bigcup_{[p] \in P} K_{[p]} \right) \cap \left(\bigcup_{[q] \in Q} K_{[q]} \right),$$

$$A_1 = \left(\bigcup_{[p] \in P} K_{[p]} \right) \setminus A, A_2 = \left(\bigcup_{[q] \in Q} K_{[q]} \right) \setminus A, B = K \setminus (A \cup A_1 \cup A_2).$$

New vertex $t+1$ and clusters from the set A create new cluster

$$[t+1] := \{t+1\} \cup A, K := (K \setminus A) \cup \{[t+1]\}.$$

A recalculation of the matrix a on renewed set of clusters K is following [2]:

$$a([t+1], [i]) := 1, [i] \in A_1 \cup \{[t+1]\}, a([i], [j]) := 1, [i] \in A_2, [j] \in A_1 \cup \{[t+1]\},$$

$$a([i], [j]) := 0, [i] \in A_1, [j] \in A_2 \cup \{[t+1]\} \cup B,$$

$$a([i], [j]) := 0, j \in A_2, [i] \in B \cup \{[t+1]\}, a([t+1], [i]) := 0, a([i], [t+1]) := 0, [i] \in B.$$

All other meanings of the matrix a elements coincide with previous ones on the step t . This algorithm has square complexity (a number of arithmetical operations) by a number of graph vertices. All other operations are assignment operations.

2 NUMERICAL EXPERIMENT AND ITS INTERPRETATION

Comparison of algorithms efficiency. To compare suggested factorization algorithms we consider the protein network Arabidopsis with 2824 vertices and 7570 edges [3]. Using the server on a base of 2 processors Xeon with 6 kernels (each of them), the frequency 2300 Hz and 32 GB of working memory it is possible to realize the factorization procedure by the calculation of the matrixes $A(k)$, $1 \leq k < t$, during 21 days. Using the sequential algorithm it is possible to fulfill the factorization procedure on the notebook with 2 kernels processor I3, the frequency 2300 Hz, and 4 GB of working memory during one and half hours.

Detection of network kernel. Results of the factorization of the network vertices are represented in Table 1 which allows to represent this protein network essentially more compact and to obtain very contrast distribution of clusters by numbers of their vertices.

numbers of cluster vertices	numbers of unisolated clusters	numbers of isolated clusters
1	1429	37
2	41	26
3	17	11
4	11	6
5	5	0
6	1	0
7	1	1
8	3	0
10	1	0
11	1	0
16	1	0
958	1	0

Table 1. Factorization of Arabidopsis protein network.

Table 1 shows that Arabidopsis protein network has a kernel with 958 vertices which composes 34 percents of all network vertices.

Analysis of kernel hubs. It is interesting to consider an interaction of kernel vertices between them and with vertices out of the kernel. Following [4], [5] we concentrate our attention on vertices which are incident with maximal numbers of edges - so called hubs. Using results of the factorization we find 10 largest numbers of edges which 1) log out from within to outside; 2) come from outside to inside; 3) log out from within to inside; 4) come from within to inside. Hubs from the set 1 may be called sources, hubs from the set 2 - sinks, hubs from the sets 3, 4 - dispatchers of the kernel. Table 2 shows that hubs from the set 1 have largest powers, hubs from the set 2 have smallest powers and hubs from the sets 3, 4 have average powers. So hubs - sources have maximal load and consequently restrict working intensity of the kernel work.

set 1	set 2	set 3	set 4
from within outside	from outside inside	from within inside	from within inside
102	23	47	50
70	19	47	39
70	16	43	38
69	14	43	37
63	12	39	33
55	8	38	27
54	7	36	27
43	6	35	26
41	5	35	24
36	4	32	24

Table 2. Numbers of edges correspond to different hubs.

Role of hubs in counteraction of kernel with its environment. Calculations show that sixty percent of edges emergent from the kernel are incident with vertices (hubs) which compose 7

percents of all appropriate kernel vertices. And sixty percents of all other types of edges are incident with kernel vertices which compose about 30 percents of all kernel vertices. This remark shows a role of hubs - sources in the considered network.

REFERENCES

1. Tsitsiashvili G.Sh. 2013. Sequential algorithms of graph nodes factorization. *Reliability: Theory and Applications*. Vol. 8(4). P. 30-33.
2. Tsitsiashvili G.Sh., Kharchenko Yu.N., Losev A.S. and Osipova M.A. 2014. Enumeration of Minimal Control Sets of Vertices in Oriented Graph. *Applied Mathematical Sciences*. Vol. 8, no. 39. P. 1941 - 1945.
3. Mingzhi Lin, Xi Zhou, Xueling Shen, Chuanzao Mao, Xin Chen. The predicted Arabidopsis Inteactome Resource and Network Topology-Based Systems Biology Analyses// *The Plant Cell*. 2011. Vol. 23. P. 911-922.
4. Liu Y.Y. , Slotine J.J., Barabasi A.L. 2011. Controllability of complex networks. *Nature*. Vol. 473. P. 167-173.
5. Barabasi L.A., Albert R.1999. Emergence of scaling in random networks. *Science*. Vol. 286. P. 509-512.