Factorization of oriented graph nodes and application to protein networks

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

Institute for Applied Mathematics FEB RAS
Vladivostok, Russia
*guram@iam.dvo.ru - www.iam.dvo.ru

Abstract

In this paper we consider oriented graph with finite sets of nodes and edges. We construct a sequential algorithm of an oriented graph factorization with a square number of arithmetical operations by a number of graph nodes to define on the graph in natural way equivalence classes of vertices (clusters) and partial order between them. A decrease of the calculation complexity is connected with an introduction of a partial order matrix which is defined recursively by assignment operations.

Method

Step 1.1 There is the single vertex 1 which creates the cluster and the set of clusters $K = \{[1]\}$.

Step 1.2 Introduce the matrix $a = \|a([p],[q])\|_{[p],[q]\in K}$ which characterizes the partial order $"\succ": a([p],[q]) = 1$, if $[p] \succ [q]$ and in opposite case $a([p],[q]) = 0$. So on the beginning we have $a([1],[1]) = 1$.

Step t.1 There is the clusters set K and the matrix a, then

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

with

$$[P] = \{[p] : t+1 \succ [p]\},\ [Q] = \{[q] : [q] \succ t+1\},$$

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

Step t.2 New vertex t + 1 and clusters from the set A create new cluster

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

and

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

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

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

Simulation Settings

We consider the protein network Arabidopsis with 2824 vertices and 7570 edges:

✓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)*, $1 \le k < t$; 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.

Simulation Results

Table. Factorization of Arabidopsis protein network.

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

Conclusions

An analysis of a distribution of equivalence classes by numbers of their nodes shows that there is the equivalence class (a kernel) with 958 nodes which composes approximately 34 percents of all network vertices.

Acknowledgements

The authors thank V.P. Bulgakov for a concession of the Arabidopsis protein network in the Cytoscape and a discussion of obtained results.