Hierarchical Classification of Directed Graph
with Cyclically Equivalent Nodes

G.Sh. Tsitsiashvili
Institute of Applied Mathematics FEB of RAS, Vladivostok, Russia
&
Far Eastern Federal University, Vladivostok, Russia

V.P. Bulgakov
Institute of Biology and Soil Science FEB of RAS, Vladivostok, Russia
&
Far Eastern Federal University, Vladivostok, Russia

A.S. Losev
Institute of Applied Mathematics FEB of RAS, Vladivostok, Russia
&
Far Eastern State Transport University, Ussuriysk, Russia

Copyright © 2016 G.Sh. Tsitsiashvili, V.P. Bulgakov and A.S. Losev. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this paper an algorithm of hierarchical classification of directed graph with cyclically equivalent nodes is constructed. An example of this algorithm application to an analysis of secondary metabolism sub-network in protein network of Arabidopsis is represented.

Mathematics Subject Classification: 05C80

Keywords: a hierarchical classification, a directed graph, cyclically equivalent nodes
Introduction

In [1], [2] an algorithm of directed graph nodes dividing into cyclically equivalent classes (clusters) is constructed. A cyclical equivalence of two nodes is an existence of a cycle containing these nodes in the directed graph. A partial order on a set of all clusters is defined and a sequential algorithm of zero-one matrix characterizing this partial order is suggested.

But this algorithm does not allow to analyse detailed a structure of single cluster which may be sufficiently complicated. Such analysis is important in some problems of the bioengineering. For example in the Arabidopsis protein network subnetworks of secondary metabolism, light, hormonal, and signal subnetworks are defined in [3].

Another approach is a hierarchical classification of undirected weighed graph $S$ nodes. This approach [4] is based on a choice of some critical value $a$ and on a transformation of an weighed graph $S$ into unweighed graph $S_a$ using a following rule. If a weight of some edge $s$ is smaller than critical value $a$ then the edge $s$ remains in the graph $S_a$, else it does not remain in $S_a$. After this transformation it is possible to find connectivity components of unweighed and undirected graph $S_a$. These components create clusters in the nodes set of the graph $S$. Successfully decreasing critical value $a$ we divide earlier obtained clusters into subclusters and so on. This algorithm allows to construct the classification of undirected weighed graph $S$ nodes. It was applied to make a hierarchical classification of plants depending on a presence or on an absence of some organic matters in them [5].

In this paper an algorithm of the hierarchical classification of the cluster nodes set is suggested. This algorithm is based on concepts of cycles with minimal length and connections between them. An idea of this algorithm is based on a procedure of the hierarchical classification of weighed undirected graph nodes.

1 Algorithm of hierarchical classification

Consider directed graph $G^1$ with nodes set $I^1$ in which all nodes are cyclically equivalent. This graph has adjacency matrix $A^1 = ||a^1_{ij}||_{i,j \in I^1}$, $a^1_{ii} = 1$ and edges with unit lengths. Using Floyd-Warshall algorithm (see [6], ”The Floyd-Warshall algorithm”, pp. 558-565 and Section 26.4, ”A general framework for solving path problems in directed graphs”, pp. 570-576.) it is possible to calculate the matrix $B^1 = ||b^1_{ij}||_{i,j \in I^1}$, $b^1_{ii} = 0$, of shortest ways lengths in the graph $G^1$. Contrast the symmetric matrix

$$C^1 = ||c^1_{ij}||_{i,j \in I^1}, \quad c^1_{ij} = c^1_{ji} = b^1_{ij} + b^1_{ji},$$
Hierarchical classification of directed graph

Considered). If in the graph $G$ clusters. Construct the matrix $A$.

2) Each two clusters $p, q \in I^2$ are cyclically equivalent in the graph $G^2$.

3) Nodes number of the graph $G^2$ is smaller than nodes number of the graph $G^1$.

Proof. 1) Assume that nodes $i, j \in p$ satisfy the equality $d^1_{ij} = 1$. Then in the graph $G^1$ there is (may be not single) cycle with the length $c^1$, which contains the nodes $i, j$. From the definition of the number $c^1$ all nodes of this cycle (of these cycles) include into the cluster $p$.

Assume now that nodes $i, j \in p$ satisfy the equality $d^1_{ij} = 0$, then it is possible to find the set of nodes $i$, $i_1$, $i_2$, ..., $i_t$, $j$, so that the equalities

$$d^1_{ii_1} = d^1_{i_1i_2} = ... = d^1_{i_{t-1}i_t} = d^1_{i_ti} = d^1_{i_{t+1}i} = ... = d^1_{ij} = 1$$

take place. Consequently in the graph $G^1$ there is the cycle $i \Rightarrow i_1 \Rightarrow ... \Rightarrow i_t \Rightarrow j \Rightarrow i_t \Rightarrow ... \Rightarrow i_1 \Rightarrow i$. All ways including this cycle include the cluster $p$ also. So in the cluster $p$ there is the cycle $i \Rightarrow i_1 \Rightarrow ... \Rightarrow i_t \Rightarrow j \Rightarrow i_t \Rightarrow ... \Rightarrow i_1 \Rightarrow i$, containing the nodes $i, j$, and all nodes of this cycle contain in the cluster $p$.

2) A cyclic equivalence of the clusters $p, q \in I^2$ in the graph $G^2$ is a corollary of a cyclic equivalence of the graph $G^1$ nodes. Indeed, assume that $i \in I^1$, $i \in p$, $j \in I^1$, $j \in q$, then in the graph $G^1$ there is a cycle containing the nodes $i, j$. This cycle passes serially through the clusters $p_0 = p, p_1, ..., p_t = q, p_{t-1}, ..., p_0 = p$. Consequently the clusters $p, q \in I^2$ are cyclically equivalent.

The statement 3) is a corollary of the number $c^1 > 1$ definition because $b^1_{i,j} + b^1_{j,i} > 1$, $i \neq j$.

Remark 1.2 1. All cycles of the graph $G^1$ with minimal length are not self intersected.
2. All nodes of a cycle with minimal length include in a same cluster.
3. An algorithm of an extraction of cycles with minimal length may be spread onto a case of a weighed graph $G$ (its edges have non unit length).

Analogously by the adjacency matrix $A^2$ and by the set $I^2$ of the directed graph $G^2$ nodes it is possible to construct the graph $G^3$ with the nodes set $I^3$ and with the adjacency matrix $A^3$. Repeat this procedure $r$ till the number of nodes in the graph $G^{r+1}$ reaches one.

As a result we obtain a hierarchical classification of the set $I^1$ nodes in the directed graph $G^1$. At each step all obtained clusters are cyclically equivalent and consist of cyclically equivalent nodes.

Another algorithm of a hierarchical classification in directed graph is based on a transformation of the matrix $C^1$ into the matrix $C^1_a = ||c^1_a(i, j)||_{(i, j) \in I^1}$ where $c^1_a(i, j)$ is a minimal length of a cycle containing nodes $i, j$. Then varying the critical value $a$ and using the algorithm of a hierarchical classification in weighed undirected graph it is possible to construct a hierarchical classification in weighed directed graph also.

2 Numerical experiment

Consider the subnetwork of secondary metabolism in the protein Arabidopsis network [3, Figure 7] (see Figure 1). Direct test shows that this subnetwork is a directed graph with cyclically equivalent nodes. Using the algorithm considered in Theorem 1.1 calculate constants defining hierarchical classification: $c^1 = 2, c^2 = 2, c^3 = 2, c^4 = 3$. On each step construct directed graphs (see Figures 2-4). On the step 4 this procedure ends.
Figure 1. Subnetwork of secondary metabolism in Arabidopsis protein network

Figure 2. Transformation of metabolism subnetwork at first step
Final result of the hierarchical classification is represented on Figure 5. A tree nodes denoted by solid circles satisfy the condition $c_i = 2$, and a node designated by an empty circle satisfies the condition $c_i = 3$. Here CIP4, HY5, COP1, COI1, JAZ3, TYFY7, MYC2, JAZ1, MYB2, TCP3, MYB113, EGL3, MYB90, PAP1, TT8, MYB114, TT2 are names of proteins containing in the secondary metabolism subnetwork.
3 Conclusion

Suggested algorithm gives the hierarchical classification of cycles in the graph. A possibility to transit to weighed graphs allows to introduce into the theoretical graph model important biological information concerning a strength of an impact of some edges on nodes (proteins) in which these edges are directed.

In Figure 5 we may see how considered subnetwork of the secondary metabolism is divided into classes of circles mainly with the length two. It is worthy to say that the proteins JAZ1 is an inhibitor and the protein TT8 is an activator which affects on the proteins MYB113, MYB114 which make the secondary metabolism products. So the inhibitor JAZ1 decelerates a work of the activator TT8 and consequently decreases an amount of the secondary metabolism products.

Figure 5. Hierarchical classification of secondary metabolism subnetwork nodes

Partially supported by Far Eastern Branch of Russian Academy Sciences, grant “Far East“ (project 15-I-4-001 o, subproject 15-I-4-030). Results of second author are supported by Russian Scientific Foundation grant (project 14-14-00230).
References


Received: June 30, 2016; Published: August 14, 2016