INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2021, VOL. 67, NO. 2, PP. 181-186
Manuscript received February 22, 2021; revised April, 2021 DOI: 10.24425/ijet.2021.135962

# Cycles in Bayesian Networks

Assem Shayakhmetova, Natalya Litvinenko, Orken Mamyrbayev,
Waldemar Wójcik, and Dusmat Zhamangarin

***Abstract*—The article is devoted to some critical problems of using Bayesian networks for solving practical problems, in which graph models contain directed cycles. The strict requirement of the acyclicity of the directed graph representing the Bayesian network does not allow to efficiently solve most of the problems that contain directed cycles. The modern theory of Bayesian networks prohibits the use of directed cycles. The requirement of acyclicity of the graph can significantly simplify the general theory of Bayesian networks, significantly simplify the development of algorithms and their implementation in program code for calculations in Bayesian networks..**

***Keywords*—Bayesian networks, directed graphs, directed cycles, propagation, Bayesian evidence**

## I. Introduction

THE theory of Bayesian Networks is nowadays widely used in different fields of science. The model built on this theory can be successfully used for a wide range of problems which contain various types of uncertainties. They find application in machine translation [1], medicine [2, 3], industry [4] or finance [5] to mention most important examples. The results of solving these problems, in most cases, are quite good and very realistically reflect reality.

Bayesian networks, as a tool for studying models with uncertainties, is considered by many authors. Pearl J. was the first one who considered more completely the Bayesian networks tool in his works [6] and [7]. Moreover, the requirement of acyclicity of directed graphs representing Bayesian networks was emphasized in these papers. The presence of cycles in directed graphs has really complicated both the theory of Bayesian networks construction and the practice of using the constructed models, although the presence of graph models with cycles was not denied. Fulfilment of the requirement of acyclicity allowed to develop many very successful software products to work with Bayesian networks (BayesiaLab, AgenaRisk, Hugin Expert etc.).

Further development of Bayesian networks can be considered as the theory of Bayesian algebraic networks, the information can be found in [8 – 10]. In this theory, the presence of directed cycles is already allowed, although many problems are still unsolved.

Problems and approaches to solving problems which models contain directed cycles are well described in various papers, for example, [9]. Even if there is one single directed cycle causes a lot of problems. The easiest option if we have several cycles that the cycles do not intersect. We have not found problems with intersecting cycles in practice or at least we have not found articles with similar models. In this paper we will consider only the simplest cases - single cycle models.

## II. Main definitions

Bayesian networks (BN) are a convenient tool for describing complex processes with various kinds of uncertainties. The Bayesian network theory is described quite well in [10, 11, 12]. Bayesian network theory is based on some sections of probability theory and graph theory [13,14]. The definitions and concepts of graph theory used in BN theory can be found in [15, 16]. The necessary concepts in probability theory can be found in [17, 18]. Features of Bayesian networks in BayesiaLab application can be found in [19].

Nevertheless, we still give some definitions from the theory of BN that are necessary for more comfortable reading of this article.

Definition 1. A graph (undirected graph) is a pair

$$G = (V(G),E(G))$$

where $E(G)$ – is a symmetric relation on the set of vertices $V(G)$, called the adjacency relation. If a given relation exists on a pair of vertices a and b of the graph, then they say that these vertices are adjacent, or that these vertices are connected by an edge. Typically, an edge is denoted by $\{a,b\}$ or $ab$. In the undirected graph $\{a,b\}=\{b,a\}$.

Def$^n$ 2. Two edges are called adjacent if they have a mutual vertex.

Def$^n$ 3. If the vertex $x$ is the end of the edge $e$, then we will say that $x$ and $e$ are incident.

Def$^n$ 4. The degree of a vertex $x$ of a graph $G$ is the number of edges that are incident to the vertex $x$. The degree of a vertex $x$ of a graph $G$ denotes as $d_G(x)$.

Def$^n$ 5. Graph with $n$ vertices is called complete and denoted as $K_n$ if any 2 vertices of this graph are connected by edge.

Def$^n$ 6. The set of vertices $U \subset V(G)$, any two of which are adjacent in the graph $G$, generate a subgraph called a clique.

Def$^n$ 7. A graph is called directed or digraph if each edge of the graph has a direction. The edge of the graph in this case is called the arc.

Def$^n$ 8. The sequence from vertex $a_0$ to vertex $a_n$ in a directed graph (in a Bayesian network) is an alternating sequence of vertices and arcs of the form

$$a_0, \{a_0 \;\; a_1\}, a_1, \{a_1 \;\; a_2\}, a_2, \{a_2 \;\; a_3\}, \cdots, a_n.$$

Def$^n$ 9. The path is a sequence without repeating arcs.

This work was supported in the framework of the grant project «Development and software implementation of a package for solving applied problems in Bayesian networks». Project URN is AP05131293

N. Litvinenko is Information and Computational Technology, 050010 Almaty, Kazakhstan (e-mail: n.litvinenko@inbox.ru).

O. Mamyrbayev and A. Shayakhmetova are with Institute of Information and Computational Technology, 050010 Almaty, Kazakhstan and Al-Farabi Kazakh National University, Almaty, Kazakhstan (e-mail: morkenj@mail.ru, asemshayakhmetova@mail.ru).

W. Wójcik is with Institute of Information and Computational Technologies CS MES RK, Almaty and Lublin Technical University, Poland (e-mail: waldemar.wojcik@pollub.pl)

Dusmat Zhamangarin is with Kazakh University Ways of Communications, Kazakhstan (e-mail: dus_man89@mail.ru)

Def$^n$ 10. A cycle is a path in which the initial and final vertices coincide.

Def$^n$ 11. The vertices *a* and *b* of a graph ***G*** are called connected if there is a path between them in the graph.

Def$^n$ 12. A graph is called connected if any two of its vertices are connected.

Def$^n$ 13. A graph is called triangular if it has no cycles without chords of length 4 or more.

Def$^n$ 14. A directed graph is called acyclic if it does not have directed cycles.

Def$^n$ 15. A Bayesian network is an acyclic directed graph with Markov condition. The vertices of the graph are often called nodes. Nodes represent some variables that reflect the main entities in the developed model. Arcs in a Bayesian network define some probabilistic connection between corresponding nodes. Sometimes such a relationship is causal. The reason is the node where the directed arc comes from, the consequence is the node where the oriented arc comes.

However, sometimes real models may contain directed cycles. Calculations in such networks are fundamentally different from calculations in ordinary Bayesian networks.

Def$^n$ 16. A skeleton of a Bayesian network is a graph obtained from a Bayesian network by replacing arcs with edges.

Def$^n$ 17. If an arc goes from the vertex *A* to the vertex *B*, then *A* is called the parent of *B*, and *B* is called the child vertex of the vertex *A*.

Def$^n$ 18. Let *Y* be some subset of vertices of a Bayesian network. *P*(*Y*) is often denoted as the set of all parents belongs to *Y*. *C*(*Y*) is often denoted as the set of all children belongs to *Y*.

Def$^n$ 19. If there is an oriented path from the vertex A to the vertex *B*, then *A* is called the ancestor of *B*, and *B* is called the descendant of *A*.

Def$^n$ 20. Two nodes are called connected if there is a sequence between them.

Def$^n$ 21. If a vertex has no ancestors, then its local probability distribution is called unconditional, otherwise conditional.

Def$^n$ 22. If the nodes are not connected by an arc, then these nodes are considering as conditionally independent.

Def$^n$ 23. Topological node numbering of a Bayesian network is a node numbering such that the number of any node is greater than the number of its parent.

Def$^n$ 24. Evidence – statements of the type "event in the node has occurred".

Def$^n$ 25. Hard evidence – statements like "an event in a node must happen".

Def$^n$ 26. Soft evidence – statements of the form "an event in a node must occur with a given probability".

Def$^n$ 27. Bayes formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Def$^n$ 28. Law of total probability. Let there is a complete set of pairwise incompatible events $A_j$. Then, for any event *B* we have the following formula for calculating its probability:

$$P(B) = \sum_{j=1}^{n} P\big(B\big|A_j\big)P(A_j)$$

Def$^n$ 29. Bayes formula (extended):

$$P(A_k B) = \frac{P(B|A_k)P(A_k)}{\sum_{j=1}^{n} P\big(B\big|A_j\big)P(A_j)}$$

## III. Cycles in Bayesian Networks

Let assume for all further reasoning, the Bayesian Networks variables (nodes) have only two states (***Y,N***). The number of states will not change our reasoning, it will only make them more complicated. To simplify further considerations, we assume that the Bayesian network contains only one directed cycle. Networks with several cycles and with intersecting cycles are a rather complicated topic and we will not consider it in this paper.

Definition. We say that the directed cycle does not contain parents if there are no vertices with parents in this cycle. I.e. there are no arcs $\{a,b\}$, where a does not belong to the cycle, but b belongs to the cycle.

Definition. We say that the directed cycle does not contain child vertices if there are no vertices with children in this cycle. I.e. there are no arcs $\{a,b\}$, where a belongs to the cycle, but b does not belong to the cycle.

If evidence is not indicated in a Bayesian network with directed cycles (hereinafter, simply in a Bayesian network), the algorithm for calculating network nodes for cycles with child nodes will not differ significantly from the case without child nodes. The presence of parents in a directed cycle significantly complicates the algorithm for calculating the vertices of a Bayesian network.

If the Bayesian network has received evidence, three cases must be distinguished:

- Vertices which do not belong to the cycle took evidences.
- Vertices which belong to the cycle took evidences.
- Both types of vertices, considered above took evidences.

In our paper we will consider the matter from easy to difficult. We will start with the simplest cases of Bayesian networks and directed cycles in the networks. Of course, not all options will be considered – the maximum volume of the article does not allow it. However, the necessary trends in the construction of algorithms can be understood. In continuation we will consider the simplest networks consisting of a single cycle and having neither parents nor child nodes. Let us state the general idea of solving these networks.

### *A. Variant 1 – Single cycle with 2 nodes*

Let us consider the simplest Bayesian Network, which contains 2 variables $A_1$ and $A_2$ (Figure 1). These 2 vertices form the single cycle. None of these vertices take Bayes evidence.


Fig. 1. Single cycle with 2 nodes

Let $X_1Y = P(A_1 = Y)$ be the probability that the variable $A_1$ takes the value $Y$, and $X_1N = 1 - X_1Y = P(A_1 = N)$ be the probability that the variable $A_1$ takes the value $N$. Conditional probabilities for arc $\{A_1, A_2\}$ and arc $\{A_2, A_1\}$ are given in

Table I and Table II respectively, where $(0 \le P_1, P_2, Q_1, Q_2 \le 1)$ are some known probabilities.

TABLE I
CONDITIONAL PROBABILITY TABLE FOR ARC $\{A_1, A_2\}$

| | $A_1 = Y$ | $A_1 = N$ |
|---|---|---|
| $A_2 = Y$ | $P_1$ | $P_2$ |
| $A_2 = N$ | $1 - P_1$ | $1 - P_2$ |

TABLE II
CONDITIONAL PROBABILITY TABLE FOR ARC $\{A_2, A_1\}$

| | $A_2 = Y$ | $A_2 = N$ |
|---|---|---|
| $A_1 = Y$ | $Q_1$ | $Q_2$ |
| $A_1 = N$ | $1 - Q_1$ | $1 - Q_2$ |

Let us calculate: $X_2Y = P(A_2 = Y)$ the probability that the variable $A_2$ takes the value $Y$, and $X_2N = 1 - X_2Y = P(A_2 = N)$ – the probability that the variable $A_2$ takes the value $N$.

$$X_2Y = X_1Y \cdot P_1 + X_1N \cdot P_2;$$
$$X_2N = X_1Y \cdot (1 - P_1) + X_1N \cdot (1 - P_2);$$

As $X_2N = 1 - X_2Y$ we can consider only the first equation.

Using the conditional probability table for the arc $\{A_2, A_1\}$ we calculate $X_1Y$ and $X_1N$.

$$X_1Y = X_2Y \cdot Q_1 + X_2N \cdot Q_2 = (X_1Y \cdot P_1 + X_1N \cdot P_2) \cdot Q_1 + \left(X_1Y \cdot (1 - P_1) + X_1N \cdot (1 - P_2)\right) \cdot Q_2;$$

$$X_1N = X_2Y \cdot (1 - Q_1) + X_2N \cdot (1 - Q_2) = (X_1Y \cdot P_1 + X_1N \cdot P_2) \cdot (1 - Q_1) + \left(X_1Y \cdot (1 - P_1) + X_1N \cdot (1 - P_2)\right) \cdot (1 - Q_2);$$

As $X_1N = 1 - X_1Y$ we can consider only the first equation.

$$X_1Y = X_1Y \cdot P_1 \cdot Q_1 + X_1N \cdot P_2 \cdot Q_1 + X_1Y \cdot Q_2 - X_1Y \cdot P_1 \cdot Q_2 + X_1N \cdot Q_2 - X_1N \cdot P_2 \cdot Q_2$$

Taking into consideration that $X_1N = 1 - X_1Y$ we obtain:

$$X_1Y = X_1Y \cdot P_1 \cdot Q_1 + (1 - X_1Y) \cdot P_2 \cdot Q_1 + X_1Y \cdot Q_2 - X_1Y \cdot P_1 \cdot Q_2 + (1 - X_1Y) \cdot Q_2 - (1 - X_1Y) \cdot P_2 \cdot Q_2$$

or

$$X_1Y = X_1Y \cdot (P_1 \cdot Q_1 - P_2 \cdot Q_1 + Q_2 - P_1 \cdot Q_2 - Q_2 + P_2 \cdot Q_2) + (Q_2 + P_2 \cdot Q_1 - P_2 \cdot Q_2)$$

or

$$X_1Y \cdot (1 - P_1 \cdot Q_1 + P_2 \cdot Q_1 + P_1 \cdot Q_2 - P_2 \cdot Q_2) = Q_2 + P_2 \cdot Q_1 - P_2 \cdot Q_2$$

or

$$X_1Y \cdot (1 - (P_1 - P_2) \cdot (Q_1 - Q_2)) = (1 - P_2) \cdot Q_2 + P_2 \cdot Q_1$$

so

$$X_1Y = \frac{(1 - P_2) \cdot Q_2 + P_2 \cdot Q_1)}{\left(1 - (P_1 - P_2) \cdot (Q_1 - Q_2)\right)} = $$
$$= \frac{(Q_2 + (Q_1 - Q_2) \cdot P_2)}{(1 - P_1 \cdot (Q_1 - Q_2) + (Q_1 - Q_2) \cdot P_2}$$

It is easy to verify that both the numerator and the denominator in this fraction are positive numbers due to the nature of the numbers $P_1$, $P_2$, $Q_1$, $Q_2$, and in addition, the denominator is greater than the numerator. I.e. $X_1Y$ satisfies the condition $0 < X_1Y < 1$. The probabilities $X_1N, X_2Y, X_2N$ are easily expressed through $X_1Y$.

### *B. Variant 2 – Single cycle with 2 nodes and evidence*

Let us consider the Bayesian Network above, which contains two variables $A_1$ and $A_2$. Let one of the vertices has obtained an evidence. If it was vertex $A_1$, then vertex $A_2$ stops to affect vertex $A_1$ and the arc $\{A_2, A_1\}$ loses its meaning, so we can stop to consider it. Vertex $A_2$ is calculated in the usual way, in accordance with conditional probability table of the arc $\{A_1, A_2\}$. If the vertex $A_2$ has obtained evidence we will have similar reasoning.

### *C. Variant 3 – Single cycle with 3 nodes*

Let us consider the simplest Bayesian Network, which contains 3 variables $A_1$, $A_2$ and $A_3$ (Figure 2). These 3 vertices form the single cycle. None of these vertices take Bayesian evidence.

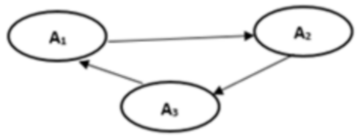


Fig. 2. Single cycle with 3 nodes.

Let $X_1Y = P(A_1 = Y)$ be the probability that the variable $A_1$ takes the value $Y$, and $X_1N = 1 - X_1Y = P(A_1 = N)$ be the probability that the variable $A_1$ takes the value $N$.

Conditional probability table for arcs $\{A_1, A_2\}$, $\{A_2, A_3\}$ and $\{A_3, A_1\}$ are given in Table III, Table IV, and Table V respectively. $(0 \le P_1, P_2, Q_1, Q_2, R_1, R_2 \le 1)$ are some known probabilities.

TABLE III
CONDITIONAL PROBABILITY TABLE FOR ARC $\{A_1, A_2\}$

| | $A_1 = Y$ | $A_1 = N$ |
|---|---|---|
| $A_2 = Y$ | $P_1$ | $P_2$ |
| $A_2 = N$ | $1 - P_1$ | $1 - P_2$ |

TABLE IV
CONDITIONAL PROBABILITY TABLE FOR ARC $\{A_2, A_3\}$

| | $A_2 = Y$ | $A_2 = N$ |
|---|---|---|
| $A_1 = Y$ | $Q_1$ | $Q_2$ |
| $A_1 = N$ | $1 - Q_1$ | $1 - Q_2$ |

TABLE V
CONDITIONAL PROBABILITY TABLE FOR ARC $\{A_3, A_1\}$

| | $A_3 = Y$ | $A_3 = N$ |
|---|---|---|
| $A_1 = Y$ | $R_1$ | $R_2$ |
| $A_1 = N$ | $1 - R_1$ | $1 - R_2$ |

Let us calculate the probability that the variable $A_2$ takes the value $Y$ as $X_2Y = P(A_2 = Y)$ and $X_2N = 1 - X_2Y = P(A_2 = N)$ – the probability that the variable $A_2$ takes the value N.

$$X_2Y = X_1Y \cdot P_1 + X_1N \cdot P_2;$$
$$X_2N = X_1Y \cdot (1 - P_1) + X_1N \cdot (1 - P_2);$$

As $X_2N = 1 - X_2Y$ we can consider only the first equation.

Using the conditional probability table for the arc $\{A_2, A_3\}$ we calculate $X_3Y$ and $X_3N$.

$X_3Y = X_2Y \cdot Q_1 + X_2N \cdot Q_2 = (X_1Y \cdot P_1 + X_1N \cdot P_2) \cdot Q_1 + (X_1Y \cdot (1 - P_1) + X_1N \cdot (1 - P_2)) \cdot Q_2;$

$X_3N = X_2Y \cdot (1 - Q_1) + X_2N \cdot (1 - Q_2) = (X_1Y \cdot P_1 + X_1N \cdot P_2) \cdot (1 - Q_1) + (X_1Y * (1 - P_1) + X_1N \cdot (1 - P_2)) \cdot (1 - Q_2);$

As $X_1N = 1 - X_1Y$ we can consider only the first equation.

$X_3Y = X_1Y \cdot (P_1 \cdot Q_1 - P_1 \cdot Q_2 + Q_2) + X_1N \cdot (P_2 \cdot Q_1 - P_2 \cdot Q_2 + Q_2)$

$X_3N = 1 - X_1Y \cdot (P_1 \cdot Q_1 - P_1 \cdot Q_2 + Q_2) - X_1N \cdot (P_2 \cdot Q_1 - P_2 \cdot Q_2 + Q_2)$

Using the conditional probability table for the arc $\{A_3, A_1\}$ we calculate $X_1Y$ and $X_1N$.

$$X_1Y = X_3Y \cdot R_1 + X_3N \cdot R_2$$

$$X_1N = X_3Y \cdot (1 - R_1) + X_3N \cdot (1 - R_2)$$

Substituting the previously found values for $X_3Y$ and for $X_3N$ into the first equation, we obtain:

$X_1Y = X_1Y \cdot (P_1 \cdot Q_1 - P_1 * Q_2 + Q_2) + X_1N \cdot (P_2 \cdot Q_1 - P_2 \cdot Q_2 + Q_2)) \cdot R_1 + (1 - X_1Y \cdot (P_1 \cdot Q_1 - P_1 \cdot Q_2 + Q_2) - X_1N \cdot (P_2 \cdot Q_1 - P_2 \cdot Q_2 + Q_2)) \cdot R_2$

$$X_1Y = \left(X_1Y \cdot \left(P_1 \cdot Q_1 + P_1 \cdot Q_2 + Q_2\right) + X_1N \cdot \left(P_2 \cdot Q_1 + P_2 \cdot Q_2 + Q_2\right)\right) \cdot R_1 + \\ + \left(1 - X_1Y \cdot \left(P_1 \cdot Q_1 - P_1 \cdot Q_2 + Q_2\right) - X_1N \cdot \left(P_2 \cdot Q_1 - P_2 \cdot Q_2 + Q_2\right)\right) \cdot R_2$$

or

$$X_1Y = \left(X_1Y \cdot \left(P_1 \cdot Q_1 + P_1 \cdot Q_2 + Q_2\right) + \left(1 - X_1Y\right) \cdot \left(P_2 \cdot Q_1 + P_2 \cdot Q_2 + Q_2\right)\right) \cdot R_1 + \\ + \left(1 - X_1Y \cdot \left(P_1 \cdot Q_1 + P_1 \cdot Q_2 + Q_2\right) + \left(1 - X_1Y\right) \cdot \left(P_2 \cdot Q_1 + P_2 \cdot Q_2 + Q_2\right)\right) \cdot R_2$$

or

$X_1Y = \left(X_1Y * (P_1 - P_2) * (Q_1 - Q_2) * (R_1 - R_2)\right) + \left((P_2 * Q_1 - P_2 * Q_2 + Q_2) * R_1 + (1 - P_2 * Q_1 + P_2 * Q_2 - Q_2) * R_2\right)$

$X_1Y * (1 - (P_1 - P_2) * (Q_1 - Q_2) * (R_1 - R_2) = P_2 * (Q_1 - Q_2) * (R_1 - R_2) + Q_2 * (R1 - R_2) + R_2$

$$X_1Y = \frac{P_2 * (Q_1 - Q_2) * (R_1 - R_2) + Q_2 * (R_1 - R_2) + R_2}{1 - (P_1 - P_2) * (Q_1 - Q_2) * (R_1 - R_2)}$$

It is easy to verify that both the numerator and the denominator in this fraction are positive numbers due to the nature of the numbers $P_1, P_2, Q_1, Q_2, R_1, R_2$, and in addition, the denominator is greater than the numerator. I.e. $X_1Y$ satisfies the condition $0 < X_1Y < 1$.

The probabilities $X_1N, X_2Y, X_2N, X_3Y, X_3N$ are easily expressed in terms of $X_1Y$.

### *D. Variant 4 – Single cycle with 3 nodes and evidence*

Let us consider the Bayesian Network above. Let one of the vertices has obtained an evidence. If it was vertex $A_2$ then vertex $A_1$ stops to affect vertex $A_2$ and the arc $\{A_1, A_2\}$ loses its meaning, so we can stop to consider it. Vertex $A_1$ is calculated in the usual way, in accordance with conditional probability table for the arc $\{A_3, A_1\}$.

### *E. Variant 5 – Single cycle with N nodes*

Let us consider the simplest Bayesian Network, which contains $N$ variables $A_1, A_2, A_3, \dots, A_N$ (Figure 3). These $N$ vertices form the single cycle.

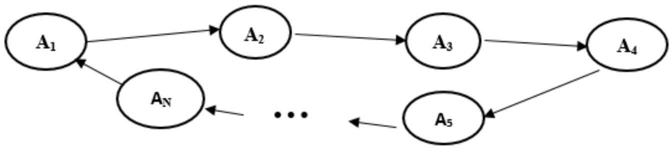


Fig. 3. Single cycle with N nodes.

None of these vertices take an evidence. In this network, each vertex has exactly one parent and exactly one child vertex. Consideration of the previous options shows that setting conditional probability tables for each arc in non-trivial cases uniquely determines the probability value at each vertex. We denote the elements of the conditional probability tables by:

$$p(A_2|A_1), p\left(A_2\middle|\overline{A_1}\right), p(A_3|A_2), p\left(xA_3\middle|\overline{A_2}\right), \dots, p(A_N|A_{N-1}),$$
$$p\left(A_N\middle|\overline{A_{N-1}}\right), p(A_1|A_N), p\left(A_1\middle|\overline{A_N}\right)$$

Arguing similarly to the previously considered variants, we can construct the following system of linear equations:

$$\begin{cases} p(A_1) = p(A_1|A_N) \cdot p(A_N) + p\left(A_1\middle|\overline{A_N}\right) \cdot (1 - p(A_N)) \\ p(A_2) = p(A_2|A_1) \cdot p(A_1) + p\left(A_2\middle|\overline{A_1}\right) \cdot (1 - p(A_1)) \\ p(A_3) = p(A_3|A_2) \cdot p(A_2) + p\left(A_3\middle|\overline{A_2}\right) \cdot (1 - p(A_2)) \\ \dots \\ p(A_N) = p(A_N|A_{N-1}) \cdot p(A_{N-1}) + p\left(A_N\middle|\overline{A_{N-1}}\right) \cdot (1 - p(A_{N-1})) \end{cases}$$

Let us denote

$$a_{kj} = p\left(A_k\middle|\overline{A_j}\right) - p\left(A_k\middle|A_j\right)$$

We obtain the following system of equations:

$$\begin{cases} p(A_1) + a_{1N} * p(A_N) = p\left(A_1\middle|\overline{A_N}\right) \\ a_{21} * p(A_1) + p(A_2) = p\left(A_2\middle|\overline{A_1}\right) \\ a_{32} * p(A_2) + p(A_3) = p\left(A_3\middle|\overline{A_2}\right) \\ a_{42} * p(A_3) + p(A_4) = p\left(A_4\middle|\overline{A_3}\right) \\ \quad \dots \\ a_{N,N-1} * p(A_{N-1}) + p(A_N) = p\left(A_N\middle|\overline{A_{N-1}}\right) \end{cases}$$

Or matrix equation:

$$\boldsymbol{A} * \boldsymbol{P} = \boldsymbol{B}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & a_{1N} \\ a_{21} & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & a_{32} & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & a_{43} & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & a_{54} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & a_{N,N-1} & 1 \end{pmatrix} * \begin{pmatrix} p(A_1) \\ p(A_2) \\ p(A_3) \\ p(A_4) \\ p(A_5) \\ \dots \\ p(A_N) \end{pmatrix} =$$

$$= \begin{pmatrix} p\left(A_1\middle|\overline{A_N}\right) \\ p\left(A_2\middle|\overline{A_1}\right) \\ p\left(A_3\middle|\overline{A_2}\right) \\ p\left(A_4\middle|\overline{A_3}\right) \\ p\left(A_4\middle|\overline{A_3}\right) \\ \dots \\ p\left(A_N\middle|\overline{A_{N-1}}\right) \end{pmatrix}$$

Determinant is equal to:

$$det\, \boldsymbol{A} = 1 + (-1)^N * a_{1N} * a_{21} * a_{32} * \dots * a_{N,N-1}$$

This determinant in invertible cases is greater than 0. Thus, there is a unique solution for this system, i.e. the probabilities of all vertices of the cycle are uniquely determined.

From examples above we can see that the values at the nodes of the cycle are uniquely determined by the conditional probability tables of arcs of this cycle. Therefore, any additional condition may cause some contradictions in the Bayesian network. For example, if one of the vertices of the cycle took an evidence, this evidence will most likely not coincide with the decision obtained in the way above. However, in the process of solving a practical problem, it is necessary to somehow resolve the contradictions. The solution to the contradictions usually consists of some simplification of the Bayesian network. If a node took a certificate, for example, it is possible to break this cycle by removing one of the arcs of the cycle. The question is: which arc of the cycle is better to remove. By removing one or another arc of the cycle, we get rid of contradictions, but the solution (the value of the probabilities in the nodes of the Bayesian network) may depend on which arc we removed. It's more natural and more convenient for calculations to remove an arc that comes to a vertex that has taken an evidence.

There are other ways to remove the contradictions. For example, to adjust the calculation of the influence of evidence on the following vertices. You can, for example, limit the number of vertices affected by the node that took an evidence, and leave the remaining values of the nodes of the cycle as they were without an evidence.

If several nodes of the cycle took evidences at the same time, then the amount of contradictions increases significantly, which means that a more substantial adjustment of the initial Bayesian network is required. There are many options for adjusting the source network, you just need to choose the correct paradigm of adjustment.

Effective and correct adjustment of the original Bayesian network can become one of the areas of machine learning in the field of Bayesian networks.

This concludes our consideration of the simplest cases when the network consists of a single cycle and we move on to more complicated examples.

### *F. Variant 6 – Single cycle with N nodes and child vertices*

In this variant we will consider the case when directed cycle has child vertices. Let us consider the simplest Bayesian Network, which contains $N$ variables $A_1, A_2, A_3, \dots, A_N$, forming the only directed cycle, as well as the vertex $B_1$. Vertex $B_1$ here is a child vertex of one of the vertices of the cycle, for example, of the vertex $A_1$ (Figure 4).

The choice of a vertex does not violate the generality of reasoning, since the vertices of the cycle can be simply renumbered as necessary.

In the absence of evidence, the presence of a child node does not affect the calculations. The calculation scheme is as follows:

- in the usual way described earlier, we calculate the vertices of the directed cycle;
- then we calculate the child vertex.

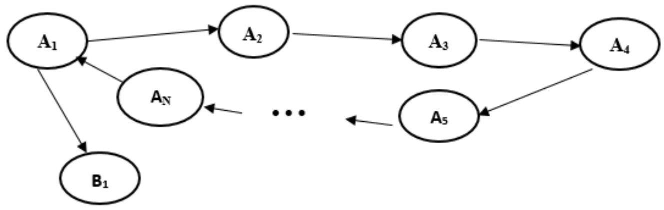


Fig. 4. Cycle with N nodes and child vertex.

There are no additional difficulties if the vertex $\boldsymbol{B_1}$ has several parents. Calculation scheme is the same. There are no additional difficulties if the directed cycle has several children vertices $\boldsymbol{B_1}, \boldsymbol{B_2}, \boldsymbol{B_3}, \dots, \boldsymbol{B_M}$. Calculation scheme is the same.

If some of the vertices of the cycle took an evidence, this also will not lead to additional difficulties, since in any case the vertices of the cycle will be calculated earlier than the child vertices.

### *G. Variant 7 – Single cycle with child vertices and evidence*

Let us consider the previous example, but when vertex $B_1$ took an evidence. The propagation should be as follows:

- Based on the evidence of vertex $B_1$ and the conditional probability table of vertex $A_1$, we calculate the vertex $A_1$.
- Based on the calculated data of the vertex $A_1$ and the conditional probability tables of the vertices $A_2$ and $A_N$, we calculate the vertices $A_2$ and $A_N$.
- Based on the calculated data of the vertex $A_2$ and the conditional probability tables of the vertices $A_3$ we calculate the vertex $A_3$. Based on the calculated data of the vertex $A_N$ and the conditional probability tables of the vertices $A_{N-1}$ we calculate the vertex $A_{N-1}$.
- Based on the calculated data of the vertex $A_3$ we calculate the vertex $A_4$. Based on the calculated data of the vertex $A_{N-1}$ we calculate the vertex $A_{N-2}$. *Etc.*

However, the values at the vertices of the directed cycle are uniquely determined by the conditional probability tables and this solution may not coincide with the solution just found. This contradiction can only be eliminated by simplifying the construction of the Bayesian network, for example by the methods described in variant 5.

If vertex $B_1$ has several parents, the amount of contradictions increases. A more significant simplification of the Bayesian network structure may be required.

If the directed cycle has several child nodes, the volume of contradictions increases even more. We need clear rules to adjust the structure of the Bayesian network in such cases.

The presence of a parent at any vertex of the cycle in the absence of evidence does not change the method for calculating the probabilities of vertices of a directed cycle. If any vertex of the cycle has several parents, the form of the matrix equation will remain the same, but the coefficients and constant terms will be calculated differently.

If several vertices of the oriented cycle have parents, the general form of the matrix will also not change.

We can make following conclusions:

- In the absence of evidence, the presence of parents at the vertices of the cycle does not change the method of calculating the probabilities of the vertices of the cycle.

- In the absence of evidence and the presence of parents at the vertices of the cycle, the probabilities of the vertices of the cycle are uniquely determined.

### *H. Variant 8 – Single cycle and parents with evidence*

Let us consider the simplest Bayesian Network, which contains the cycle with $N$ vertices $A_1, A_2, A_3, \dots, A_N$, that has $M$ parents $B_1, B_2, B_3, \dots, B_M$ (Figure 6). Let some parents take evidences.

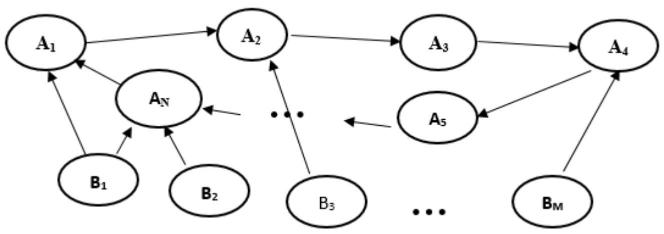


Fig. 6. Cycle with N nodes and M parent vertex

The arc between the parent vertex and the vertex (vertices) of the directed cycle means the dependence of the vertex (vertices) of the cycle on the parent vertex. Moreover, for a given arc (or arcs) it does not matter how the value was obtained at the parent vertex. This can be:

- Marginal probability of a parent vertex.
- Parent vertex probability can be calculated in previous steps.
- Parent vertex took a certificate.

Therefore, the method for calculating the vertices of a directed cycle will remain the same as in variant 8. In this case, the number of parent vertices of the directed cycle does not matter, does the vertex have one or more parents [3], is the parent vertex the parent of one or more vertices of the directed cycle. The only important that the parent vertices of the cycle are calculated before calculating the vertices in the cycle.

## IV. CONCLUSIONS

In the process of developing models of real processes using Bayesian networks, sometimes it becomes necessary to include a directed cycle in the network. The classic Bayesian network theory forbids the use of directed cycles. However, the rejection of directed cycles can sometimes lead to unnecessary simplifications of the model. In the theory of Bayesian algebraic networks, the authors have already considered the possibility of using directed cycles. However, the theory of Bayesian algebraic networks is rather fundamental and difficult to use as a model for solving practical problems.

This paper is considering the possibility of using simple directed cycles in Bayesian networks. We consider and analyzed 8 variants, covering the main ways of entering a directed cycle in a Bayesian network and methods for calculating the probabilities of the cycle vertices.

## REFERENCES

[1] A. Nafalski and A.P. Wibawa, "Machine translation with javanese speech levels' classification," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, vol. 6, no 1, pp 21-25, 2016. https://doi.org/10.5604/20830157.1194260

[2] Z.Omiotek and P. Prokop, "The construction of the feature vector in the diagnosis of sarcoidosis based on the fractal analysis of CT chest images," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, vol. 9, no. 2, pp. 16-23, 2019. https://doi.org/10.5604/01.3001.0013.2541

[3] A. Litvinenko, O. Mamyrbayev, N. Litvinenko, A. Shayakhmetova, "Application of Bayesian networks for estimation of individual psychological characteristics," *Przeglad Elektrotechniczny*, vol. 95, no. 5, pp. 92-97, 2019

[4] X.Q. Cai, X.Y. Wu, X. Zhou, "Stochastic scheduling subject to breakdown-repeat breakdowns with incomplete information," *Operations Research*, vol. 57, no. 5, pp. 1236–1249, 2009. doi:10.1287/opre.1080.0660

[5] K.W. Fornalski, "The Tadpole Bayesian Model for Detecting Trend Changes in Financial Quotations," *R&R Journal of Statistics and Mathematical Sciences*, vol. 2, no. 1, pp. 117–122, 2016.

[6] J. Pearl "Artificial Intelligence Applications", in *How to Do with Probabilities what people say you can't,*/ Editor Weisbin C.R., IEEE, North Holland, pp. 6–12, 1985.

[7] J. Pearl "Probabilistic Reasoning in Intelligent Systems". San Francisco: Morgan Kaufmann Publishers, 1988,

[8] A. Tulupiev "Algebraic Bayesian networks," in *"Logical-probabilistic approach to modeling knowledge bases with uncertainty*," SPb.: SPIIRAS, 2000.

[9] S. Nikolenko, A. Tulupiev "The simplest cycles in Bayesian networks: Probability distribution and the possibility of its contradictory assignment," SPIIRAS. Edition 2, 2004. vol.1.

[10] F.V. Jensen, T.D. Nielsen "Bayesian Networks and Decision Graphs," Springer, 2007.

[11] D. Barber, "Bayesian Reasoning and Machine Learning," 2017, 686 p. http://web4.cs.ucl.ac.uk/ staff/D.Barber/ textbook/020217.pdf

[12] R.E. Neapolitan "Learning Bayesian Networks," 704p. http://www.cs.technion.ac.il/~dang/books/Learning%20Bayesian%20Networks(Neapolitan,%20Richard).pdf

[13] O. Mamyrbayev, M. Turdalyuly, N. Mekebayev, and et al. "Continuous speech recognition of kazakh language», AMCSE 2018 Int. conf. On Applied Mathematics, Computational Science and Systems Engineering, Rom, Italy, 2019, vol. 24, pp. 1-6.

[14] A. Litvinenko, N. Litvinenko, O. Mamyrbayev, A. Shayakhmetova, M. Turdalyuly "Clusterization by the K-means method when K is unknown," Inter. conf. Applied Mathematics, Computational Science and Systems Engineering. Rome, Italy, 2019, vol. 24, pp. 1-6.

[15] O.Ore "Graph theory," Moscow: Science, 1980, 336 p.

[16] Ph. Kharari "Graph theory," Moscow: Mir, 1973, 300 p.

[17] V. Gmurman "Theory of Probability and Mathematical Statistics: Tutorial," Moscow: 2003, 479 p.

[18] A.N. Kolmogorov "Theory: Manual," in *"Basic Concepts of Probability,"* Moscow: Science, 1974.

[19] N. Litvinenko, A. Litvinenko, O. Mamyrbayev, A. Shayakhmetova "Work with Bayesian Networks in BAYESIALAB," Almaty: IPIC, 2018, 311 p. (in Rus). ISBN 978-601-332-206-3.