PONTIFICAL CATHOLIC UNIVERSITY OF MINAS GERAIS
POSTGRADUATE PROGRAM IN COMPUTER SCIENCE

João Pedro Santos Pereira

Increasing descriptive capacity for longitudinal studies analysis through triadic rules

Belo Horizonte

September 23, 2024

João Pedro Santos Pereira

Increasing descriptive capacity for longitudinal studies analysis through triadic rules

Dissertation submitted to the Postgraduate Program in Computer Science at the Pontifical Catholic University of Minas Gerais, as a partial requirement for the degree of Master in Computer Science.

Advisor: Prof. Dr. Luis Enrique Zárate Gálvez

Belo Horizonte

September 23, 2024

FICHA CATALOGRÁFICA
Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

Pereira, João Pedro Santos
P436i Increasing descriptive capacity for longitudinal studies analysis through triadic rules / João Pedro Santos Pereira. Belo Horizonte, 2024.
49 f. : il.

Orientador: Luis Enrique Zárate Gálvez
Dissertação (Mestrado) - Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Informática

1. Mineração de dados (Computação). 2. Estruturas de dados (Computação). 3. Conceitos - Análise. 4. Percepção de padrões. 5. Banco de dados - Análise. 6. Saúde. 7. Doenças - Evolução - Análise. I. Gálvez, Luis Enrique Zárate. II. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Informática. III. Título.

CDU: 681.3.011

Ficha catalográfica elaborada por Daniela Luzia da Silva Gomes - CRB 6/2505

João Pedro Santos Pereira

Increasing descriptive capacity for longitudinal studies analysis through triadic rules

Dissertation submitted to the Postgraduate Program in Computer Science at the Pontifical Catholic University of Minas Gerais, as a partial requirement for the degree of Master in Computer Science.

Prof. Luis Enrique Zárate – PUC Minas

Prof. Rokia Missaoui – Université du Québec en Outaouais (UQO), Canadá

Prof. Gilmar Reis – PUC Minas

Prof. Mark A. Junho Song – PUC Minas

Belo Horizonte, September 23, 2024.

ACKNOWLEDGMENT

The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (Grant PROAP 88887.842889/2023-00 – PUC/MG, Grant PDPG 88887.708960/2022-00 – PUC/MG - INFORMÁTICA and Finance Code 001).

RESUMO

Um tipo comum de banco de dados usado no campo médico é a base de dados longitudinal, que registra observações clínicas ao longo do tempo para uma mesma amostra de pacientes, denominadas ondas. Este tipo de base de dados fornece informações importantes para quem deseja analisar a evolução de doenças, observando mudanças ao longo do tempo em uma população específica.

Neste trabalho, a abordagem adotada para analisar esse tipo de dado é a Análise de Conceitos Triádicos, que permite a obtenção de regras triádicas, correspondentes a regras de associação com condições, para descrever as relações temporais presentes nas bases de dados.

Para demonstração da técnica, dois casos de estudo foram elaborados. O primeiro utilizando uma base de dados a respeito de pacientes com HIV e a síndrome do olho-seco, e o segundo caso de estudo apresenta uma base de dados de uma pesquisa telefônica em relação a pandemia do COVID-19.

Os resultados obtidos demonstram que a Análise de Conceitos Triádicos é uma estratégia promissora para descrever bancos de dados longitudinais. Além disso, no contexto dos estudos de caso mencionados, os resultados destacaram a possibilidade de aprimorar sistemas de suporte à decisão clínica, adaptando-os aos tratamentos específicos de cada doença.

Essa abordagem tem o potencial de fornecer percepções valiosas para profissionais de saúde, permitindo uma compreensão mais abrangente das relações temporais entre sintomas e tratamentos, e auxiliando na melhoria dos protocolos de tratamento e na tomada de decisões clínicas mais embasadas.

Palavras-chave: Mineração de dados. Mineração de dados longitudinais. Análise de Conceito Triádico. Regras Triádicas. Saúde.

ABSTRACT

A common type of database used in the medical field is the longitudinal database, which records clinical observations over time for the same sample of patients, referred to as waves. This type of database provides crucial information for analyzing disease progression, observing changes over time within a specific population.

In this work, the approach adopted to analyze this type of data is the Triadic Concept Analysis, which allows the extraction of triadic rules, corresponding to association rules with conditions, to describe the temporal relationships present in the databases.

To demonstrate the technique, two case studies were conducted. The first utilized a database concerning patients with HIV and dry eye syndrome, while the second case study involved a telephone survey database related to the COVID-19 pandemic.

The obtained results demonstrate that Triadic Concept Analysis is a promising strategy for describing longitudinal databases. Furthermore, in the context of the mentioned case studies, the results highlighted the possibility of enhancing clinical decision support systems by adapting them to the specific treatments of each disease.

This approach has the potential to provide valuable insights for healthcare professionals, enabling a comprehensive understanding of the temporal relationships between symptoms and treatments, and assisting in improving treatment protocols and making more informed clinical decisions.

Keywords: Data mining. Longitudinal data mining. Triadic Concept Analysis. Triadic Rules. Health.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Longitudinal studies are a widely used approach in the health field, involving the collection of observations over time from the same sample of individuals, recorded in distinct periods called waves (DIGGLE, 1994). These studies are of great importance as they allow for the tracking of clinical evolution and individuals' responses over time. For instance, a longitudinal database might contain information on patients' clinical, psychological, or environmental conditions before, during, and after a specific treatment. These databases are a valuable resource for describing, analyzing, and validating procedures, as well as extracting useful and non-obvious insights that support clinical decision-making (RIBEIRO et al., 2017).

However, the dimensionality of the data can vary in longitudinal databases. In some studies, such as those involving human aging in a population, there can be hundreds of variables and hundreds of thousands of records (RIBEIRO et al., 2017). In other cases, like studies on the efficacy of a vaccine, the number of variables might be reduced, with only a few dozen variables and a few thousand records. In studies on the treatment of rare or specific diseases, it is common to have a limited number of variables, typically a few dozen, with a reduced number of records, often less than a hundred (CHAUDHURI et al., 2007), (GOTTAPU; DAGLI, 2018), (SHAMIR et al., 2017), (WILKINSON et al., 2019), and (BALNE et al., 2017a).

This variation in data dimensionality can impact the applicability of data mining and machine learning techniques. In databases with a large number of records, it is possible to search for patterns and build models with predictive and generalization capabilities. Conversely, in databases with a small number of records, descriptive analyses are more suitable for extracting insights from the available data.

Regardless of the number of available records, descriptive analyses can be performed on longitudinal databases. These analyses aim to associate variables of interest, such as clinical data within the same wave or across different waves.

In this work, we propose the use of Formal Concept Analysis (FCA) theory (GANTER; WILLE, 2012), an area of applied mathematics based on the theory of conceptual lattices, to analyze longitudinal study databases. FCA aims to represent and extract knowledge from datasets involving objects (individuals), attributes (clinical conditions, symptoms, etc.), and their incidence relationships. These elements can be represented through a dyadic formal context, from which association rules between attributes can be extracted (AGRAWAL; IMIELIńSKI; SWAMI, 1993). Triadic Concept Analysis (TCA), proposed by (LEHMANN; WILLE., 1995), is an extension of FCA that uses a triadic formal context, introducing a third element called condition, which determines the incidence relationship between objects and attributes.

In the healthcare context, the objects correspond to patients, the attributes to symptoms, and the conditions to the different waves of the longitudinal study. From the triadic context, triadic association rules can be extracted. In this work, two types of rules will be used: Condition-Based Association Rules (BACAR) and Attribute-Based Association Rules (BCAAR). These rules allow us to observe the relationships between conditions for certain attributes (BACAR) and the relationships between attributes across conditions (BCAAR). With these rules, it is possible, for example, to observe relationships between waves based on recurring symptoms and to explore the relationships between symptoms in a specific wave (or across waves).

Although FCA has received significant attention in the field of Data Mining in recent years (SINGH; KUMAR; GANI, 2016a), the application of triadic theory remains quite limited and is entirely absent in the analytical description of longitudinal studies. Therefore, describing the clinical/symptomatic characteristics of individuals undergoing a longitudinal study for the analysis of clinical treatments can be crucial for understanding the effects of the procedures adopted during treatment. In this work, we present an approach to extract triadic rules in appropriately modeled triadic contexts to describe the relationship between attributes and waves. The results obtained from the case study analyzing the HIV viral load in patients with dry eye symptoms demonstrate the potential of TCA to describe and analyze longitudinal study databases in the healthcare field.

This work is organized as follows: Section 2 presents a review of related work; Section 3 provides the theoretical background for longitudinal databases and triadic formal concept analysis; Section 4 describes the methodology developed for database analysis using TCA; Section 5 presents the application and analysis of the results in the considered case studies; Section 6 discusses the implications of the findings; and finally, conclusions and future directions are presented in Section 7.

1.1 Motivation

Information is essential in a world dominated by technology. In recent years, technological advancements have led to the discovery and creation of various data generation tools. However, this vast amount of raw data needs to be analyzed and interpreted to be transformed into useful information and knowledge. One such type of data is what we call longitudinal data, which results from tracking the same group over time. These data are crucial for understanding patterns and trends that can be applied in various fields, from public health to consumer behavior and financial markets.

In this context, there arises an opportunity to use machine learning tools to analyze this type of data and gain valuable insights. Through this kind of analysis, it is possible to improve the accuracy of predictions and the effectiveness of interventions, whether in disease treatment or in political or financial situations.

Therefore, the main motivation of this research is to investigate how machine learning techniques, with an emphasis on triadic analysis, can be applied to longitudinal databases to extract knowledge.

1.2 Objectives

This work aims to explore the applicability of FCA methods when applied to longitudinal databases. Despite the challenges and limitations associated with these types of data, we aim to show that it is possible to obtain information and improve the prediction of temporal trends through this analysis.

1.2.1 General objective

Investigate how FCA methods can be applied to longitudinal data to extract insights and better describe databases, including case studies presenting real-world scenarios.

1.2.2 Specific objectives

- To review the currently most used techniques for longitudinal data analysis.
- To propose a model for conducting triadic analysis on longitudinal data.
- To apply this model to real-world scenarios of longitudinal data.
- To evaluate and interpret the results of the conducted case studies.
- To demonstrate the practical applicability of this analysis in different contexts.

2 RELATED WORK

2.1 Longitudinal studies

Longitudinal studies are a rich source of information, containing data that tracks the evolution of a group of people over time. Because of this, the scientific community has made efforts to explore this type of database.

In the healthcare field, a longitudinal study is a valuable type of data. By analyzing observations over a period of time, it is possible to monitor the evolution of a group of patients and gather important information for decision-making in a healthcare setting. With that in mind, the authors of the article (LORENZO et al., 2023) conducted a longitudinal study to investigate the mental health responses of Canadian university students during the COVID-19 pandemic, focusing on anxiety, depression, and perceived stress.

A longitudinal study can also be very useful when attempting to compare and associate different conditions, whether between groups or within a group of individuals. The study (GUGUSHVILI; JAROSZ, 2024) investigates the association between perceived social position and subsequent health-related quality of life in the Polish population from 2008 to 2018.

Longitudinal studies have been widely used in researching a wide range of diseases in the literature, including cancer (HU et al., 2019), depression (ROTENSTEIN et al., 2016), diabetes (LI et al., 2022), Parkinson's disease (YASSINE et al., 2022), tuberculosis (SETIANINGRUM et al., 2022), and human aging (RIBEIRO et al., 2017). In all of these studies, longitudinal data provide valuable information about the characteristics of each disease, such as its progression, treatment outcomes, and prevention strategies.

Regarding the techniques used in longitudinal studies, most research applies methods such as linear regression, logistic regression, and multilevel logistic regression, conducting statistical analyses of the datasets (RIBEIRO et al., 2017). These statistical methods can help understand how various factors may change over time and influence outcomes, providing insights and detecting predictive factors.

However, there is a lack of effort towards a more in-depth analysis of this type of database. The use of machine learning techniques to analyze longitudinal data can offer valuable insights and complement traditional statistical methods. In this article, we aim to analyze longitudinal data using machine learning techniques, specifically association rules through triadic analysis, to gain information, discover hidden patterns, and better describe this type of database.

2.2 FCA for knowledge extraction

Since its formalization, Formal Concept Analysis (FCA) has been continuously developed and applied in various domains, such as Information Retrieval (KUMAR; RADVANSKY; ANNAPURNA, 2012), Security Analysis (POELMANS; ELZINGA; DEDENE, 2013), Text Mining (MAIO et al., 2014), Topic Detection (CIGARRÁN; CASTELLANOS; GARCÍA-SERRANO, 2016), Social Network Analysis (MISSAOUI; KUZNETSOV; OBIEDKOV, 2017), Search Engines (NEGM; ABDELRAHMAN; BAHGAT, 2017), among others (SINGH; KUMAR; GANI, 2016b). For instance, in the healthcare field, FCA has been used to identify redundancies among different clinical tests and propose substitutions with more cost-effective alternatives (GUPTA; KUMAR; BHATNAGAR, 2007).

Formal Concept Analysis (FCA) is an approach that allows for efficient data analysis and representation of knowledge by identifying associations and dependencies through a binary incidence relationship between objects and attributes (DAVENPORT; PRUSAK, 1997; GANTER; STUMME; WILLE, 2005). Several scientific articles have explored the use of FCA in association rule extraction, enabling the identification of patterns and relationships among different variables in datasets (STUMME; WILLE; WILLE, 1998; BASTIDE et al., 2000; WILLE, 2001; STUMME et al., 2002).

However, many healthcare datasets exhibit more complex relationships, involving ternary and even n-ary relationships (BAZIN, 2020). Consequently, there has been a growing interest in proposing solutions for the analysis and exploration of these multidimensional data, especially in triadic contexts (FELDE; STUMME, 2021). Some studies have compared different triadic algorithms, such as TRICONS, TRIAS, and DATA-PEELER, to identify the most suitable ones for specific contexts (TRABELSI; JELASSI; YAHIA, 2012). Moreover, strategies have been proposed for discovering optimal patterns in triadic datasets (IGNATOV et al., 2015).

To address the challenge of analyzing triadic data, some researchers have explored transforming triadic contexts into equivalent dyadic contexts, enabling the application of traditional FCA methods (SELMANE et al., 2013). Other studies have compared specific algorithms for triadic context analysis to identify the most effective approaches (ZHUK; IGNATOV; KONSTANTINOVA, 2014). Additionally, tools like Lattice-Miner have been developed for extracting triadic association rules (MISSAOUI; EMAMIRAD, 2017) and implications (GANTER; OBIEDKOV, 2004).

The articles (LANA et al., 2022) and (NORONHA et al., 2022) are likely the first works to explore triadic analysis for describing longitudinal studies in the health field. The first article addresses the effectiveness of COVID-19 infection prevention methods, while the second study investigates patterns related to human aging by observing the evolution of clinical and environmental conditions over time in individuals.

The objective of this work is to reinforce the potential of triadic analysis for examining longitudinal databases through the extraction of triadic association rules based on conditions. These rules can be used to analyze longitudinal data, incorporating the temporal aspect inherent in this type of study, and contribute to enhance clinical decision support systems in the treatment of various diseases.

3 BACKGROUND

3.1 Longitudinal databases

Longitudinal studies in the health field are designed to track individuals or groups over time, recording a variety of information relevant to healthcare. This information can include clinical data such as medical test results, diagnoses, and treatments; symptomatic data such as the presence and severity of symptoms; psychological and emotional data such as stress levels and quality of life; environmental data such as exposure to certain agents or conditions; and other types of data that may be pertinent to the study at hand.

Depending on the study design, it is possible to include new individuals in the study population over time. This approach can be used, for instance, to analyze the impact of specific events or interventions on a larger sample. Additionally, researchers can introduce new variables of interest as the study progresses. This flexibility enables researchers to explore various aspects and correlations within the collected data more comprehensively.

Generically, a longitudinal database can be represented by the following equation (RIBEIRO; ZáRATE, 2019):

$$V = \begin{bmatrix} v_{1111} & \cdots & v_{111F_{11}} & v_{1121} & \cdots & v_{112F_{12}} & \cdots & v_{11C_1 1} & \cdots & v_{11C_1F_{1C_1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{1R_111} & \cdots & v_{1R_11F_{11}} & v_{1R_121} & \cdots & v_{1R_12F_{12}} & \cdots & v_{1R_1C_11} & \cdots & v_{1R_1C_1F_{1C_1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{N111} & \cdots & v_{N11F_{11}} & v_{N121} & \cdots & v_{N12F_{12}} & \cdots & v_{N1C_N1} & \cdots & v_{N1C_NF_{NC_N}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{NR_N11} & \cdots & v_{NR_N1F_{11}} & v_{NR_N21} & \cdots & v_{NR_N2F_{12}} & \cdots & v_{NR_NC_N1} & \cdots & v_{NR_NC_NF_{NC_N}} \end{bmatrix} \tag{3.1}$$

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_N \end{bmatrix} \begin{matrix} R_1 \times \sum_{n=1}^{C_1} |F_{1n}| \\ \vdots \\ R_N \times \sum_{n=1}^{C_N} |F_{Nn}| \end{matrix}$$

$$for \quad 1 \leq i \leq N; \quad 1 \leq j \leq |R_i|\,; \quad 1 \leq k \leq |C_i|\,; \quad 1 \leq l \leq |F_{ik}|$$

where each element $v_{i,j,k,l}$ corresponds to a data point in wave i, record j, variable category k, and variable l within category k. N corresponds to the number of waves in the dataset, $|R_i|$ corresponds to the number of records in wave i, $|C_i|$ corresponds to the number of variable categories in wave i, and $|F_{ik}|$ corresponds to the number of variables in wave i within category k.

Note that the number of records, categories, and variables can vary between waves in longitudinal studies, as they may involve the inclusion of new participants or the collection of different types of data at different times. Therefore, longitudinal studies in health represent a valuable approach to better understand changes and correlations over time. They provide important insights for the healthcare field and contribute to advancing medical and scientific knowledge.

3.2 Formal Concept Analysis

Formal Concept Analysis (FCA) is a branch of applied mathematics based on the theory of conceptual lattices (GANTER; WILLE, 2012). FCA identifies formal concepts from the incidence relationship between objects and attributes present in a dataset.

The primary definition of FCA is the formal dyadic context, which corresponds to a tuple $K := (G, M, I)$, where G corresponds to a set of objects, M a set of attributes, and I indicates an incidence relation ($I \subseteq \mathrm{G} \times M$), describing which objects have specific attributes. The incidence relation can also be represented by (G,M) ∈ I, which can be understood as "OBJECT G HAS ATTRIBUTE M", and in the healthcare context "PATIENT G HAS SYMPTOM M". Table 1 presents an example of dyadic context.

Table 1 – Dyadic context example

G/M	m1	m2	m3
g1	x		x
g2		x	
g3		x	x

Given a formal context $K := (G, M, I)$ and a subset of objects $A \subseteq G$, it might be interesting to know which attributes from M are common to all objects A. Similarly, it can also be interesting to determine which objects from G share the same attributes from $B \subseteq M$. At this point, the derivation operator (') is used, which is defined by Equations 3.2 and 3.3:

$$A' = \{m \in M \mid gIm \; \forall \; g \in A\} \tag{3.2}$$

$$B' = \{g \in G \mid gIm \; \forall \; m \in B\} \tag{3.3}$$

A formal concept extracted from a formal context $K := (G, M, I)$ is defined by a pair (A, B), where A corresponds to a subset of objects, $A \subseteq G$ (*extension*), and B corresponds to a subset of attributes, $B \subseteq M$ (*intent*), with $A' = B$ and $B' = A$.

It is also possible to extract implication rules from the formal context or the set of formal concepts. Thus, an implication is of the form $P \to Q$, where P and Q are sets of attributes and $P' \subseteq Q'$. In other words, every object that has the attributes of P also has the attributes of Q.

3.3 Triadic Concept Analysis

For some problems, such as the one considered in this work, it is convenient to identify implication rules associated with a condition. In the theory of Triadic Concept Analysis (TCA), introduced by (LEHMANN; WILLE., 1995), the triadic context is formally defined as a quadruple $\mathcal{K} = (K1, K2, K3, Y)$, where $K1$, $K2$, and $K3$ are sets, and Y corresponds to a ternary incidence relation between $K1$, $K2$, and $K3$. In other words, $Y \subseteq K1 \times K2 \times K3$, where the elements of $K1$, $K2$, and $K3$ are called objects, attributes, and conditions, respectively.

This relation can be represented by $(g, m, b) \in Y$, which can be read as: "the object g has the attribute m under the condition b". An example of triadic context is shown in Table 2, where we can consider the patients as objects, each wave as a condition, and the symptoms as attributes. Thus, it is possible to represent the patients with their symptoms/clinical observations for each study wave.

Table 2 – Triadic context example

	Wave1		**Wave2**		**Wave3**	
Patient	Symptom1	Symptom2	Symptom1	Symptom2	Symptom1	Symptom2
1	X			X		X
2	X		X	X		
3		X	X	X	X	

In TCA, a triadic formal context can be transformed into a diadic formal context with restrictions, where an incidence between object (patient) and attribute (clinical condition, symptom) in a specific condition (wave in the longitudinal study) cannot be exclusive to other symptoms for the same wave (WILLE, 2009). Table 3 presents the transformation of the triadic formal context, given in Table 2, to the equivalent dyadic formal context.

Table 3 – Example of transformation from a triadic context to a dyadic context

Patient	Symptom1-W1	Symptom2-W1	Symptom1-W2	Symptom2-W2	Symptom1-W3	Symptom2-W3
1	X			X		X
2	X		X	X		
3		X	X	X	X	

This transformation, resulting in a formal dyadic context, may be necessary depending on the objective and tools used for analysis. However, in some cases, this conversion is unnecessary, since there are currently tools that can work directly with the triadic context.

3.4 Association rules

According to the literature, (BIEDERMANN, 1997) was the first work to address the problem of extracting implication rules from a triadic context. From these contexts, it is possible to extract two types of rules that will be used in this work. These rules are referred to as *Biedermann Conditional Attribute Association Rule* (BCAAR) and *Biedermann Attributional Condition Association Rule* (BACAR) (MISSAOUI; KWUIDA., 2011), which are formalized as follows, considering a triadic formal context $\mathcal{K} = (K1, K2, K3, Y)$:

- **BACAR**: $(C1 \rightarrow C2)\ A\ (sup,\ conf)$, where $C1,\ C2 \subseteq K3$ and $A \subseteq K2$, that is, when $C1$ occurs for all attributes in A, then $C2$ will also occur for the same attributes with support (sup) and confidence ($conf$).
- **BCAAR**: $(A1 \rightarrow A2)\ C\ (sup,\ conf)$, where $A1,\ A2 \subseteq K2$ and $C \subseteq K3$, that is, when $A1$ occurs under all conditions in C, then $A2$ will also occur under the same conditions with support (sup) and confidence ($conf$).

In this work, we use the metrics of *Support* and *Confidence* to evaluate triadic association rules. Considering $P,\ Q$ as sets of variables of interest (attributes) for a triadic formal context $\mathcal{K} = (K1, K2, K3, Y)$. For the association rule $P \rightarrow Q$, extracted from the context $\mathcal{K}$, the metrics are defined as:

- **Support**: The support metric corresponds to the proportion of objects in the subset $g \in G$ that satisfy the implication $P \rightarrow Q$, relative to the total number of objects $\mid G \mid$ in the formal context $\mathcal{K}$, as shown in Equation 3.4, where $(.)'$ corresponds to the derivation operator defined in Equation 3.3.

$$Support\ (P \rightarrow Q) = \frac{\mid (P \cup \{Q\})' \mid}{\mid G \mid} \tag{3.4}$$

It is important to note that the support value should not be solely considered to compare whether an implication is more relevant than another (a common interpretation when considering frequent item mining). The support metric measures the proportion of cases of variables of interest in the association rule for a particular wave or between waves. For instance, when the support value is high, the relationship between variables is a recurrent observation in the study population. On the other hand, when the value is low, it may indicate that the relationship is an exception and should not be disregarded from a clinical point of view.

- **Confidence**: The confidence metric corresponds to the proportion of objects $g \in G$ that contain P and also contain Q, relative to the total number of objects $\mid G \mid$, as shown in Equation 3.5.

$$Confidence\ (P \rightarrow Q) = \frac{\mid (P \cup \{Q\})' \mid}{\mid P' \mid} \tag{3.5}$$

4 PROPOSED METHOD FOR TRIADIC ANALYSIS

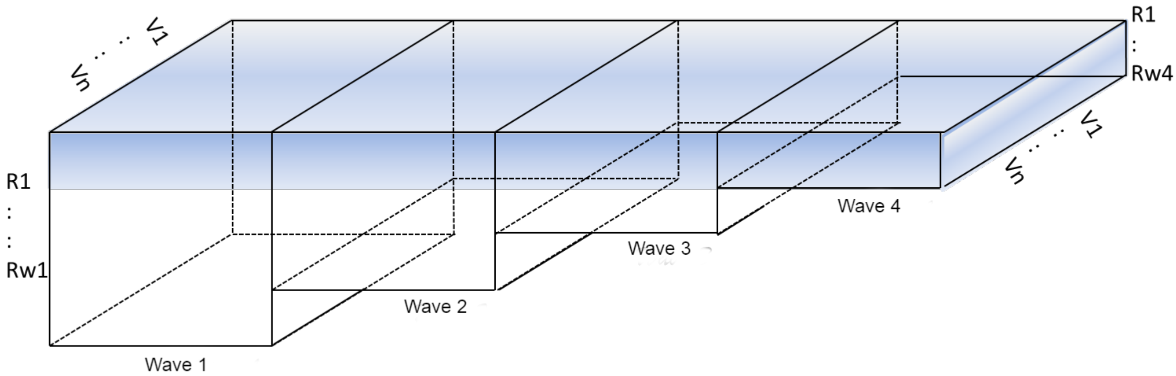
This section presents the necessary procedures for applying triadic analysis to longitudinal study databases.

4.1 Restrictions for longitudinal databases

To enable the analysis, the following restrictions must be placed on the database:

1. The number of records present in each wave must be the same for all waves considered in the triadic analysis. That is, $|R_i| = |R_{i+1}|$ for $i = 1..N-1$, where N corresponds to the number of waves considered in the study.
2. The records (sample individuals) must be the same for all waves considered, that is, $ID_{patient}R_i = ID_{patient}R_{i+1}$ for $i = 1...N-1$.
3. The number of variable categories in each wave must be the same for all waves considered in the triadic analysis, that is, $|C_i| = |C_{i+1}|$ for $i = 1...N-1$.
4. The variable category in each wave must be the same for all waves considered in the triadic analysis, that is, $CategoryC_i = CategoryC_{i+1}$ for $i = 1...N-1$.
5. The number of variables per category must be the same for all waves considered in the triadic analysis, that is, $|F_{ik}| = |F_{i+1k}|$ for $i = 1...N-1$ and $k = 1...|C_i|$.
6. The variables per category must be the same for all waves considered in the triadic analysis, that is, $VariablesF_{ik} = VariablesF_{i+1k}$ for $i = 1...N-1$ and $k = 1...|C_i|$.

Figure 1 – Longitudinal dataset illustration



After applying the presented restrictions, the dataset considered for triadic analysis must have the same records, the same variables (clinical observations/symptoms), and no missing data for all the analyzed waves, as shown in Figure 1.

4.2 Characterization of triadic analysis

It is important to emphasize that triadic analysis is not a classification task. If the considered records (individuals/patients) characterize distinct populations, the triadic analysis must be applied to one of the specific populations. Thus, triadic analysis describes the population of a class of records through association rules and concepts. For example, when the objective is analyzing clinical records of the evolution of patients who underwent a specific treatment, one must first remove the records of patients who received a placebo treatment and work only with the population one wishes to describe.

Regarding the variables of interest, those that represent the population under analysis should be selected. For this, it is important to have knowledge from domain experts who can assist in choosing these variables.

4.3 Defining the triadic relation

To indicate the binary incidence of the relationship between objects and attributes (patients and variables), it is important to consider thresholds as reference values for each variable of interest. Values above a threshold may indicate either favorable or unfavorable clinical conditions for the patient concerning a disease.

In this work, it is considered that incidence relationships should occur for either favorable or unfavorable conditions regarding a health state, but not the mixture of both conditions within the same triadic formal context. As said in the previous Subsection 4.2, it is necessary to take only one specific population, in this case, take only the favorable or unfavorable conditions. This restriction arises because the algorithms for extracting association rules, within the theory of formal concept analysis, are not capable of distinguishing health situations when both conditions are considered within the context.

4.4 Triadic association rule extraction

Finally, after constructing the triadic context, this context must be converted into a file format suitable for the selected tool to perform the extraction of triadic association rules.

At this stage, it may be necessary to convert the triadic context into an equivalent dyadic context, depending on the requirements of the selected tool, as previously mentioned in Subsection 3.2.

4.5 Lattice Miner

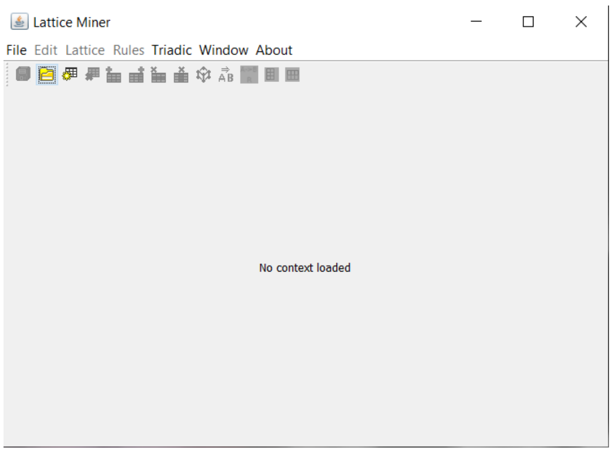
Lattice Miner is a data mining application developed by the LARIM research laboratory at the Université du Québec en Outaouais, under the supervision of Professor Rokia Missaoui. This tool enables the generation of formal concepts and association rules, with its primary objective being to assist in the visualization and exploration of pattern discovery.

4.5.1 Usage steps

1) Home screen

Program's home screen, where the first information are displayed. By selecting *Triadic > Open*, the file explorer will open, allowing to select the desired database file for analysis. For demonstration purposes, we will use one of the various test datasets provided by the program's authors, *lehmann_wille*.

Figure 2 – Home screen



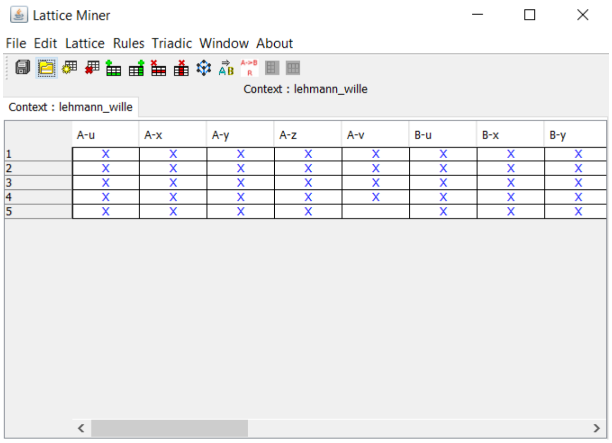
2) Loaded context

Once the context is loaded, the program displays the objects listed on the left, the attributes across the top, and incidences are represented by an X between an object and an attribute. Note that the context conditions are associated with the attributes; in this example, the values A and B are attributes, while the values u, x, y, z and v are conditions.

3) Defining support and confidence metrics

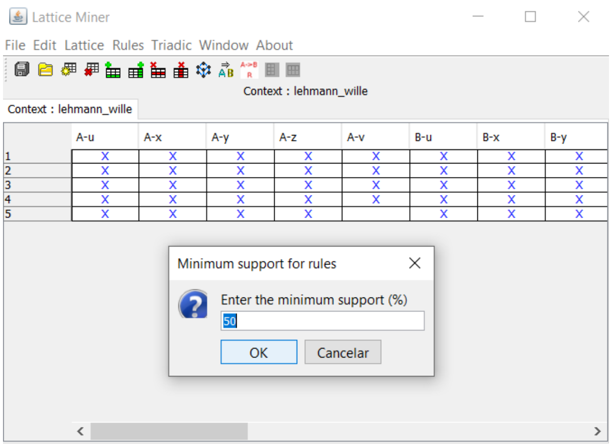
After an initial visualization of the database, it is time to generate the association rules for the context. By selecting *Triadic > BACARs* (or *BCAARs*), the following window will open. In this screen, we set the minimum values for Support and Confidence

Figure 3 – Loaded context



for the analysis. It is important to remember that the lower these values, the more rules will be generated, resulting in more work to identify the most interesting ones.

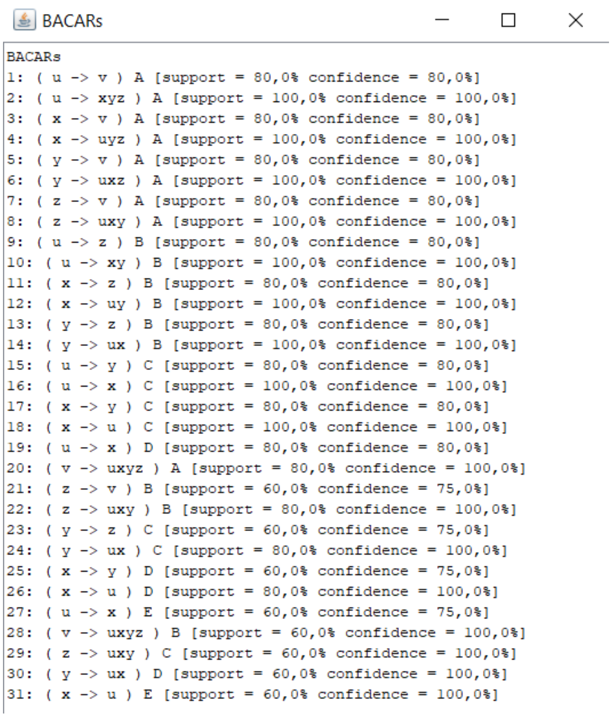
Figure 4 – Defining support and confidence metrics



4) Generated rules

Finally, the last screen displays all the rules generated with the specified Support and Confidence levels. As a final task, it is crucial to select and interpret the rules that are most relevant to the problem.

Figure 5 – Generated rules

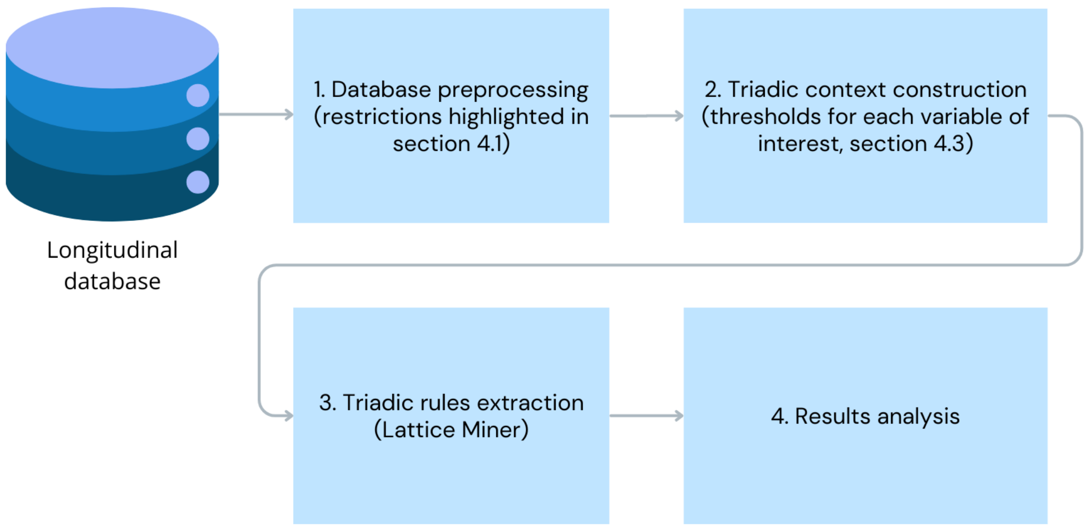


```
BACARs
1: ( u -> v ) A [support = 80,0% confidence = 80,0%]
2: ( u -> xyz ) A [support = 100,0% confidence = 100,0%]
3: ( x -> v ) A [support = 80,0% confidence = 80,0%]
4: ( x -> uyz ) A [support = 100,0% confidence = 100,0%]
5: ( y -> v ) A [support = 80,0% confidence = 80,0%]
6: ( y -> uxz ) A [support = 100,0% confidence = 100,0%]
7: ( z -> v ) A [support = 80,0% confidence = 80,0%]
8: ( z -> uxy ) A [support = 100,0% confidence = 100,0%]
9: ( u -> z ) B [support = 80,0% confidence = 80,0%]
10: ( u -> xy ) B [support = 100,0% confidence = 100,0%]
11: ( x -> z ) B [support = 80,0% confidence = 80,0%]
12: ( x -> uy ) B [support = 100,0% confidence = 100,0%]
13: ( y -> z ) B [support = 80,0% confidence = 80,0%]
14: ( y -> ux ) B [support = 100,0% confidence = 100,0%]
15: ( u -> y ) C [support = 80,0% confidence = 80,0%]
16: ( u -> x ) C [support = 100,0% confidence = 100,0%]
17: ( x -> y ) C [support = 80,0% confidence = 80,0%]
18: ( x -> u ) C [support = 100,0% confidence = 100,0%]
19: ( u -> x ) D [support = 80,0% confidence = 80,0%]
20: ( v -> uxyz ) A [support = 80,0% confidence = 100,0%]
21: ( z -> v ) B [support = 60,0% confidence = 75,0%]
22: ( z -> uxy ) B [support = 80,0% confidence = 100,0%]
23: ( y -> z ) C [support = 60,0% confidence = 75,0%]
24: ( y -> ux ) C [support = 80,0% confidence = 100,0%]
25: ( x -> y ) D [support = 60,0% confidence = 75,0%]
26: ( x -> u ) D [support = 80,0% confidence = 100,0%]
27: ( u -> x ) E [support = 60,0% confidence = 75,0%]
28: ( v -> uxyz ) B [support = 60,0% confidence = 100,0%]
29: ( z -> uxy ) C [support = 60,0% confidence = 100,0%]
30: ( y -> ux ) D [support = 60,0% confidence = 100,0%]
31: ( x -> u ) E [support = 60,0% confidence = 100,0%]
```

5 CASE STUDIES

In this section, the conducted case studies are presented. Figure 6 illustrates the stages used in each case study during the application of triadic analysis.

Figure 6 – Methods description for triadic analysis



5.1 Case 1: HIV and dry eye syndrome

This case study applies triadic analysis to a longitudinal database of patients with Human Immunodeficiency Virus (HIV) and dry eye syndrome (BALNE et al., 2017b). In (AGRAWAL et al., 2016), the authors demonstrated that over half of the individuals with HIV also presented some eye-related disease. Therefore, it becomes important to study the correlation between both conditions.

Further information about dry eye syndrome can be found in (LEMP; FOULKS, 2007), where the authors discuss its differences from other eye disorders and classify this syndrome based on the severity of each case.

5.1.1 Materials

The database used for this case study comes from a longitudinal study of patients with HIV and dry eye syndrome during HAART therapy (SHAFER; VUITTON, 1999).

The longitudinal study was conducted with thirty-four patients over three medical appointments, defined as "visits" in this case study, each one with a three-month interval. During each of these visits, the following procedures were performed: Schirmer's test on both eyes (to assess the level of eye dryness) and a blood test (to measure the percentage of CD4 and CD8 lymphocytes, as well as the HIV viral load in the patient's blood).

5.1.2 Methods description

1) Preprocessing

The database contains records of patients who were absent due to their non-participation in one or two visits, making it necessary to remove them from the database (restriction [2] of our methodology, see Section 4). This procedure resulted in a total of sixteen patients being removed. Considering that the second visit had the highest amount of missing data, it was also removed from the database. Altogether, sixteen patients remained at the base.

After removing some demographic data, the following variables were considered for analysis: results of Schirmer tests in both eyes, CD4 and CD8 cell counts in the blood, and the result of HIV viral load detection in the blood.

2) Triadic context construction

This study used reference values to determine whether the examination results showed abnormality or not. For the Schirmer test results, any value below or equal 10 mm/5min was considered abnormal (KARAMPATAKIS et al., 2010). For the CD4 and CD8 cell counts in the blood examination, values below 500 cells/μL for CD4 and values above 1000 cells/μL for CD8 were considered abnormal (BOFILL et al., 1992).

When constructing the triadic context, patients correspond to objects, while the results of clinical exams are considered as attributes, and visits are treated as conditions. Table 4 presents the variables, study conditions, and the criteria for considering a value abnormal in parentheses (reference values).

Table 4 – Variables and conditions for the "HIV and dry eye syndrome" case study

Variable	**Variable description (reference values)**
SchirmerRight	Schirmer Test in the right eye ($\leq$ 10 mm/5 min)
SchirmerLeft	Schirmer Test in the left eye ($\leq$ 10 mm/5 min)
CD4	CD4 cell count ($<$ 500 cells/μL)
CD8	CD8 cell count ($>$ 1000 cells/μL)
ViralLoad	HIV viral load (detected)
Condition	**Condition description**
Visit1	First visit for measurement of values
Visit3	Third visit for measurement of values

3) Triadic rules extraction

After constructing the triadic context and applying the Lattice Miner tool, 10 BACAR-type and 62 BCAAR-type rules were generated. Association rules with minimum support and confidence of 6.25% were selected so that all the rules reflected at least one record in the database.

The choice of a minimum support of 6.2% was considered relevant because the records covered by this threshold represent an exception compared to the more frequent patterns. Since triadic analysis aims to describe the longitudinal database, individuals with low support should be observed as exceptions to the rule. They may provide interesting insights that should be further investigated in greater depth.

Table 5 presents the main BACAR-type association rules generated. These rules relate the conditions (clinical examination visits) that share the same attributes (clinical measurement results). In other words, these association rules link the visits that have similar examination results, allowing us to analyze which tests consistently showed abnormal results over time.

For example, rule number 1 (BACAR) shows that whenever in the first visit patients have a blood test with a CD4 cell count below 500 cells/μL, then they also do so in the third visit, with support of 93.8% and confidence of 100%.

Table 5 – BACAR rules for the "HIV and dry eye syndrome" case study

N	**Rule**	**Support**	**Confidence**
1	(Visit1 $\rightarrow$ Visit3) CD4	93,8%	100,0%
2	(Visit1 $\rightarrow$ Visit3) ViralLoad	37,5%	60,0%
3	(Visit1 $\rightarrow$ Visit3) SchirmerLeft	18,8%	42,9%
4	(Visit1 $\rightarrow$ Visit3) CD8	31,2%	100,0%
5	(Visit1 $\rightarrow$ Visit3) SchirmerRight, ViralLoad	12,5%	66,7%
6	(Visit1 $\rightarrow$ Visit3) CD4, CD8	25,0%	100,0%
7	(Visit1 $\rightarrow$ Visit3) SchirmerRight, SchirmerLeft, ViralLoad	6,2%	50,0%

Table 6 presents the main BCAAR-type association rules generated. These rules relate the attributes (clinical results) that are present under the same conditions (visits). In this context, these association rules link the abnormal test results within the same visits, allowing us to analyze the relationships between different clinical tests during each visit.

For example, rule number 12 (BCAAR) indicates that whenever patients exhibited abnormalities in Schirmer's tests in both the right and left eyes in the first visit, then they also had a blood test with detectable HIV viral load during the first visit, with support of 12.5% and confidence of 40%.

Table 6 – BCAAR rules for the ”HIV and dry eye syndrome” case study

N	Rule	Support	Confidence
1	(SchirmerLeft → CD4) Visit3	50,0%	100,0%
2	(SchirmerRight → CD8) Visit3	31,2%	62,5%
3	(SchirmerRight → CD4) Visit3	50,0%	100,0%
4	(SchirmerLeft → ViralLoad) Visit1	18,8%	42,9%
5	(SchirmerLeft → CD8) Visit1	12,5%	28,6%
6	(SchirmerLeft → CD4) Visit1	43,8%	100,0%
7	(SchirmerRight → ViralLoad) Visit1	18,8%	50,0%
8	(SchirmerRight → CD4) Visit1	37,5%	100,0%
9	(SchirmerRight, SchirmerLeft → CD8) Visit3	25,0%	66,7%
10	(SchirmerRight, SchirmerLeft → CD4) Visit3	37,5%	100,0%
11	(SchirmerLeft → CD4) Visit1, Visit3	18,8%	100,0%
12	(SchirmerRight, SchirmerLeft → ViralLoad) Visit1	12,5%	40,0%
13	(SchirmerRight, SchirmerLeft → CD8) Visit1	6,2%	20,0%
14	(SchirmerRight, SchirmerLeft → CD4) Visit1	31,2%	100,0%
15	(SchirmerRight → CD4, ViralLoad) Visit1, Visit3	12,5%	100,0%

4) Results analysis

We can highlight some of the BCAAR rules, as these indicate relationships between different clinical tests. Rules 4, 7, 12, and 15 show patients who had abnormalities in Schirmer's Test and also had a blood test with a detectable HIV viral load. However, these are not the only rules that can relate these two conditions. Rules 1, 2, 3, 5, 6, 8, 9, 10, 11, 13, and 14 involve patients who showed abnormalities in Schirmer's Test and also had abnormal levels of CD4 and CD8 in their blood, which are components of the immune system and the main targets of the virus. Therefore, these association rules can assist in linking dry eye syndrome to HIV.

5.2 Case 2: COVID-19

This case study consists of applying Triadic Analysis to a longitudinal database regarding a telephone interview conducted in Belgium during the COVID-19 pandemic to analyze the psychological behavior of a population (De Coninck et al., 2022), (De Coninck; D'HAENENS; MATTHIJS, 2020).

5.2.1 Materials

The database used comes from a survey conducted through telephone interviews with residents of Belgium during the early months of the COVID-19 pandemic.

The longitudinal study involved one thousand six hundred and forty-six participants and was divided into five interviews, starting on March 17, 2020, and ending on April 5, 2021. The interview includes questions regarding the following aspects of individuals: sociodemographic data, perceived vulnerability to the disease and attitudes towards vaccines, social media consumption, informal care, physical activity, contact with the coronavirus, attitudes towards public health measures and government, and finally, personality traits.

5.2.2 Methods description

1) Preprocessing

The longitudinal dataset does not have missing data regarding the three hundred and forty-eight participants who took part in all five waves of the study; therefore, no removal or missing data treatment was necessary.

The following interview questions were considered for the analysis: "PVD11 - My hands feel dirty after touching money", "PVD12 - I am likely to catch a cold, flu, or other illness, especially if it is 'going around'", "PVD13 - It makes me anxious to be around sick people", "government_crisis - Respondent thinks the government is handling the crisis well".

2) Triadic context construction

The answers to the interview questions are numerical values between 1 and 7, with 1 being "completely disagree" and 7 being "completely agree." For discretization purposes, it was considered that a response of 5, 6, or 7 indicates that the individual agrees with the statement, while a response of 1, 2, or 3 indicates that the individual does not agree.

When constructing the triadic context, the interviewees correspond to the objects, while the interview questions are considered attributes, and finally, the interview waves are treated as conditions. Table 7 presents the variables and conditions of the study.

3) Triadic rules extraction

After constructing the triadic context and applying Lattice Miner, 287 BACAR-type rules and 134 BCAAR-type rules were generated.

Table 8 presents the main BACAR-type association rules generated. These rules relate the conditions (waves of interviews) that share the same attributes (interview responses). In other words, these association rules connect the waves of interviews that have similar responses, allowing us to analyze how many individuals maintained the same responses over time.

For example, rule number 9 (BACAR) indicates that whenever in Wave1

Table 7 – Variables and conditions for the ”COVID-19” case study

Variable	**Variable description**
PVD11	”My hands feel dirty after touching money”
PVD12	”I am likely to catch a cold, flu, or other illness, especially if it is ’going around’”
PVD13	”It makes me anxious to be around sick people”
government_crisis	Respondent thinks government is handling crisis well
Condition	**Condition description**
Wave1	First wave of interviews, from March 17 to March 22, 2020
Wave2	Second wave of interviews, from April 6 to April 18, 2020
Wave3	Third wave of interviews, from May 17 to June 5, 2020
Wave4	Fourth wave of interviews, from August 18 to August 31, 2020
Wave5	Fifth wave of interviews, from March 17 to April 5, 2021

Table 8 – BACAR rules for the ”COVID-19” case study

N	**Rule**	**Support**	**Confidence**
1	(Wave1 → Wave2) PVD11	24,7%	64,7%
2	(Wave2 → Wave3) PVD11	27,3%	67,9%
3	(Wave3 → Wave4) PVD11	24,1%	63,2%
4	(Wave4 → Wave5) PVD11	25,3%	67,2%
5	(Wave1 → Wave2) PVD12	35,3%	69,1%
6	(Wave2 → Wave3) PVD12	31,0%	59,0%
7	(Wave3 → Wave4) PVD12	28,7%	67,1%
8	(Wave4 → Wave5) PVD12	30,5%	68,8%
9	(Wave1 → Wave2) PVD13	28,4%	76,2%
10	(Wave2 → Wave3) PVD13	30,2%	60,0%
11	(Wave3 → Wave4) PVD13	29,3%	68,9%
12	(Wave4 → Wave5) PVD13	26,1%	62,8%
13	(Wave1 → Wave2) government_crisis	50,0%	82,5%
14	(Wave2 → Wave3) government_crisis	35,6%	56,4%
15	(Wave3 → Wave4) government_crisis	18,7%	47,4%
16	(Wave4 → Wave5) government_crisis	16,1%	67,5%

respondents state that they feel anxious when around sick people they also do so in Wave2, with support of 28.4% and confidence of 76.2%.

Table 9 presents the main BCAAR-type association rules generated. These rules relate the attributes (interview responses) that are present under the same conditions (waves of interviews). In this context, these association rules link the responses to different questions within the same waves of interviews.

For example, rule number 1 (BCAAR) indicates that whenever the interviewed individuals reported feeling susceptible to diseases, especially if it is ”going around” then they also reported feeling anxious around sick people, in the first interview wave, with

Table 9 – BCAAR rules for the "COVID-19" case study

N	Rule	Support	Confidence
1	(PVD12 → PVD13) Wave1	23,3%	45,5%
2	(PVD12 → PVD13) Wave2	31,6%	60,1%
3	(PVD12 → PVD13) Wave3	26,7%	62,4%
4	(PVD12 → PVD13) Wave4	24,7%	55,8%
5	(PVD12 → PVD13) Wave5	24,1%	52,8%
6	(PVD12 → government_crisis) Wave1	29,3%	57,3%
7	(PVD12 → government_crisis) Wave2	32,2%	61,2%
8	(PVD12 → government_crisis) Wave3	18,1%	42,3%
9	(PVD12 → government_crisis) Wave4	11,5%	26,0%
10	(PVD12 → government_crisis) Wave5	15,8%	34,6%
11	(PVD12, PVD13 → government_crisis) Wave1	13,8%	59,3%
12	(PVD12, PVD13 → government_crisis) Wave2	19,3%	60,9%
13	(PVD12, PVD13 → government_crisis) Wave3	11,5%	43,0%
14	(PVD12, PVD13 → government_crisis) Wave4	6,0%	24,4%
15	(PVD12, PVD13 → government_crisis) Wave5	8,9%	36,9%

support of 23.3% and confidence of 45.5%.

4) Results analysis

Analyzing BACAR-type rules allows us to track associations between attributes across waves. In this case study, we can observe how the responses to an interview question changed or remained consistent over the interview waves. For example, rules 13, 14, 15, and 16 (BACAR) show the evolution of the variable "government_crisis" from the first to the last wave. Through the confidence values, it is possible to see that 82.5% of respondents who believed the government was handling the crisis well in the first wave continued to believe so in the second wave. This value decreased to 56.4% between the second and third waves, decreased again to 47.4% between the third and fourth waves, and increased to 67.5% between the fourth and last wave.

These values can be insightful as they reflect the population's opinion regarding governmental measures concerning the pandemic. By tracking how these opinions evolve across the interview waves, the government can gain a better understanding of public sentiment and adapt in the formulation of more effective and well-targeted policies to address the challenges posed by the pandemic.

On the other hand, BCAAR-type rules show the relationship between different attributes in the same set of waves. In this case study, we can observe how interviewed individuals agreed with different statements in the same interview waves. For example, rules 1, 2, 3, 4, and 5 (BCAAR) show the relationship between the variables "PVD12" and "PVD13" across interview waves. Through the support values, it is possible to perceive that 23.3% of respondents in the first wave felt susceptible to diseases and also felt anxious around sick people. This value increased to 31.6% in the second wave, decreased to 26.7% in the third wave, decreased again to 24.7% in the fourth wave, and finally ended at 24.1% in the last wave.

It is interesting to note that in BCAAR-type rules, the support value increased in the second wave of interviews. This could suggest a heightened level of concern or awareness among the population during that period, which may have been caused by various factors, such as an increase in COVID-19 cases, changes in governmental policies, media coverage of the pandemic, or personal experiences with the virus.

6 DISCUSSION AND IMPLICATIONS

The case studies showed the potential use of triadic analysis to describe longitudinal studies, particularly in disease treatments and clinical patient monitoring. Through triadic rules, this analysis can extract valuable knowledge and insights from longitudinal databases, enhancing decision-making support. These insights are manifested as association rules among clinical factors or variables derived from the database. Such rules can suggest new hypotheses, which can be subsequently tested and validated through targeted clinical studies.

Triadic rules, accompanied by support and confidence measurements, offer insights into the clinical profiles of patients' reactions before, during, and after treatment, as well as associations between different clinical symptoms throughout the treatment. Despite being applied to small datasets initially, Triadic Concept Analysis (TCA) has shown promise in enhancing the analysis of longitudinal databases. With larger datasets, this technique enables the extraction of more representative association rules, thereby improving predictive capabilities.

Triadic analysis can identify a substantial number of rules by thoroughly exploring the formal triadic context defined by the problem domain. This extensive exploration opens the door to numerous new clinical studies. The number of generated rules can be adjusted by setting specific support and confidence thresholds relevant to the study. In machine learning, association rules with high support and confidence levels are typically valued and considered strong (ALAWADH, 2022), such as in Market Basket Analysis, greater than 60%. The decision on which rules should be kept and which should be discarded during the mining process is based on the values of these two indices.

From a computational point of view, when the support/confidence index decrease, the number of rules generated increases significantly, making the analysis process difficult for the user. Experiments presented in (ZHENG; KOHAVI; MASON, 2001) demonstrate that mining with real databases can lead to thousands of association rules with high computational effort (AGRAWAL; SRIKANT, 1994). However, association rules with lower support values (less than 60%, 50% or 30%) should not be dismissed as irrelevant in the health area. As long as there is evidence in the database, these rules might indicate significant cases of exceptions and should also be examined.

The triadic analyses presented in this work allow obtaining the following insights: for patients with comorbidities (case 1), exploring the cross-sectional bonds between two or more symptoms of distinct comorbidities during a specific wave of the longitudinal study, and examining the persistence of one or more symptoms of distinct comorbidities over time through longitudinal observation. For a population experiencing extreme health conditions, it is possible to evaluate the psychological effects (case 2), analyzing the cross-sectional associations between emotional feelings within a specific wave, and understanding how emotional feelings persist across different waves of the study.

Furthermore, the study's findings also have several managerial implications. Primarily, they highlight the importance of extracting insights from longitudinal studies to enhance decision-making and diagnostic processes. By facilitating the early identification of risk patterns, these insights can significantly improve clinical outcomes. That is, the association rules generated can reveal significant patterns that contribute to promptly identifying risks. This early detection enables quicker and more effective interventions, thereby enriching treatment success.

Applying Triadic Concept Analysis to longitudinal health databases offers significant benefits for both healthcare professionals and health services managers. For professionals, TCA identifies factors associated with the health of specific populations, enabling adjustments in clinical strategies for more effective medical interventions over time. For health service managers, longitudinal observation of groups via TCA allows exploration of factors influencing the emergence of chronic diseases. This insight supports proactive preventive measures, including targeted public policy interventions, which can mitigate long-term healthcare costs by addressing the root causes of chronic conditions earlier. Thus, TCA embraces not only the effectiveness of clinical practices but also aids in strategic planning and cost management in healthcare.

7 CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This work shows the potential of using triadic analysis to describe longitudinal studies in the healthcare field for decision-making support. Through this analysis, we can extract insights from a longitudinal database, highlighting the relationships between factors or variables within the database. These insights can suggest new hypotheses that could be verified through more specific and evidence-based studies.

The products of this analysis are expressed using association rules derived with support and confidence measures. These rules can, for instance, provide a better understanding of the clinical profile of patients before, during, and after a given treatment, as well as the relationships between different clinical symptoms throughout the treatment.

Despite being conducted on a reduced dataset, it is worth noting that triadic analysis has proven promising for improving decision-making and analyzing longitudinal databases. Important applications of this technique can be identified, such as the ability to be applied to larger datasets, more accurately describe the patients in the study, and predict the future outcomes of a treatment.

It is important to highlight that the proposed method cannot analyze two distinct populations of diseases with similar symptoms, as established in Section 4.2. Triadic association rules would relate only to the symptoms, and we would not have a clear separation between the diseases. However, what can be done is to analyze a single population that presents two or more distinct diseases, each with its distinct symptoms. This is demonstrated in the first case study conducted, where we analyzed a population of patients affected by HIV and dry eye syndrome and obtained rules relating the distinct symptoms.

Note that triadic analysis can identify a large number of rules (since this analysis fully explores the constructed formal triadic context), opening the possibility for various new studies. The number of generated rules can be adjusted by the support and confidence measures of interest for the study. In machine learning, association rules that exhibit high levels of support and confidence are typically valued and considered strong (ALAWADH, 2022) (for example, in Market Basket Analysis, greater than 60%). The decision about which rules should be retained and which should be discarded during the mining process is based on the values of these two indices.

From a computational standpoint, when the support/confidence indices decrease, the number of generated rules increases significantly, making the analysis process more challenging for the user. Experiments presented in (ZHENG; KOHAVI; MASON, 2001) show that mining with real databases can lead to thousands of association rules with high computational effort (AGRAWAL; SRIKANT, 1994).

In the healthcare field, however, association rules with support values considered low (less than 60%, 50%, or 30%) should not be dismissed as irrelevant, provided there is evidence in the database, making them exceptions that should also be studied. As an example, studies show that some unvaccinated individuals exposed to Covid-19 did not develop the disease or produce antibodies but had more numerous and effective defense cells (ASTBURY et al., 2022).

Considering the computational aspect, finding triadic association rules from triadic contexts can require substantial computational effort. Depending on the sparsity degree of the context, the computation time can reach several hours of processing. However, this work aimed to show the potential of triadic analysis to describe longitudinal databases.

7.2 Future Work

For future work, it would be interesting to perform an analysis using different degrees of severity for each variable in each case study. This analysis would allow for the observation of differences in the severity of each symptom and its progression throughout the treatment.

Since this is a multidisciplinary work, the assistance of professionals in the field relevant to the problem domain is greatly appreciated. Their help is crucial in the process of selecting the most critical variables, as well as the most valuable rules for the study.

Additionally, triadic analysis can be applied to various other domains, such as in nutrition (monitoring individuals during a diet or training), epidemiological control (tracking individuals over time after vaccination), etc. We believe that triadic analysis can be applied to any longitudinal database that meets the criteria described in the methodology.

REFERENCES

AGRAWAL, R. et al. A distinct cytokines profile in tear film of dry eye disease (ded) patients with hiv infection. Cytokine, Elsevier, v. 88, p. 77–84, 2016.

AGRAWAL, R.; IMIELIńSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. New York, NY, USA: Association for Computing Machinery, 1993. (SIGMOD '93), p. 207–216. ISBN 0897915925. Disponível em: <https://doi.org/10.1145/170035.170072>.

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994. (VLDB '94), p. 487–499. ISBN 1558601538.

ALAWADH, A. M. B. M. M. A survey on methods and applications of intelligent market basket analysis based on association rule. Journal on Big Data, v. 4, n. 1, p. 1–25, 2022. ISSN 2579-0056. Disponível em: <http://www.techscience.com/jbd/v4n1/47664>.

ASTBURY, S. et al. Hla-dr polymorphism in sars-cov-2 infection and susceptibility to symptomatic covid-19. Immunology, v. 166, n. 1, p. 68–77, 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/imm.13450>.

BALNE, P. K. et al. Dataset of longitudinal analysis of tear cytokine levels, cd4, cd8 counts and hiv viral load in dry eye patients with hiv infection. Data in Brief, v. 11, p. 152–154, 2017. ISSN 2352-3409. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352340917300148>.

BALNE, P. K. et al. Dataset of longitudinal analysis of tear cytokine levels, cd4, cd8 counts and hiv viral load in dry eye patients with hiv infection. Data Brief, v. 11, n. C, p. 152–154, 2017.

BASTIDE, Y. et al. Mining frequent patterns with counting inference. SIGKDD Exploration Newsletter, p. 66–75, 2000.

BAZIN, A. On implication bases in n-lattices. Discrete Applied Mathematics, v. 273, p. 21–29, 2020. ISSN 0166-218X. Advances in Formal Concept Analysis: Traces of CLA 2016.

BIEDERMANN, K. How triadic diagrams represent conceptual structures. In: LUKOSE, D. et al. (Ed.). Conceptual Structures: Fulfilling Peirce's Dream. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997. p. 304–317. ISBN 978-3-540-69424-3.

BOFILL, M. et al. Laboratory control values for cd4 and cd8 t lymphocytes. implications for hiv-1 diagnosis. Clinical & Experimental Immunology, Wiley Online Library, v. 88, n. 2, p. 243–252, 1992.

CHAUDHURI, K. R. et al. The metric properties of a novel non-motor symptoms scale for parkinson's disease: results from an international pilot study. MOVEMENT DISORDERS, Wiley Online Library, v. 22, n. 13, p. 1901–1911, 2007.

CIGARRÁN, J.; CASTELLANOS, Á.; GARCÍA-SERRANO, A. A step forward for Topic Detection in Twitter: An FCA-based approach. EXPERT SYSTEMS WITH APPLICATIONS, v. 57, p. 21–36, sep 2016. ISSN 09574174.

DAVENPORT, T. H.; PRUSAK, L. Working knowledge: How organizations manage what they know. In: HARVARD BUSINESS SCHOOL PRESS. [S.l.], 1997.

De Coninck, D.; D'HAENENS, L.; MATTHIJS, K. Perceptions and opinions on the covid-19 pandemic in flanders, belgium: Data from a three-wave longitudinal study. DATA IN BRIEF, v. 32, p. 106060, 2020. ISSN 2352-3409. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352340920309549>.

De Coninck, D. et al. Updating 'perceptions and opinions on the covid-19 pandemic in flanders, belgium' with data of two additional waves of a longitudinal study. DATA IN BRIEF, v. 42, p. 108010, 2022. ISSN 2352-3409. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352340922002219>.

DIGGLE, P. J. Analysis of longitudinal data. TECHNOMETRICS, v. 45, p. 181 – 181, 1994.

FELDE, M.; STUMME, G. Triadic exploration and exploration with multiple experts. In: BRAUD, A. et al. (Ed.). FORMAL CONCEPT ANALYSIS - 16TH INTERNATIONAL CONFERENCE, ICFCA 2021, STRASBOURG, FRANCE, JUNE 29 - JULY 2, 2021, PROCEEDINGS. [S.l.]: Springer, 2021. (Lecture Notes in Computer Science, v. 12733), p. 175–191.

GANTER, B.; OBIEDKOV, S. Implications in triadic formal contexts. ICCS, p. 186–195, 2004.

GANTER, B.; STUMME, G.; WILLE, R. Formal concept analysis: foundations and applications. In: VOL. 3626, SPRINGER SCIENCE & BUSINESS MEDIA. [S.l.], 2005.

GANTER, B.; WILLE, R. FORMAL CONCEPT ANALYSIS: MATHEMATICAL FOUNDATIONS. [S.l.]: Springer Science & Business Media, 2012.

GOTTAPU, R. D.; DAGLI, C. H. Analysis of parkinson's disease data. PROCEDIA COMPUTER SCIENCE, Elsevier, v. 140, p. 334–341, 2018.

GUGUSHVILI, A.; JAROSZ, E. A longitudinal study of perceived social position and health-related quality of life. SOCIAL SCIENCE MEDICINE, v. 340, p. 116446, 2024. ISSN 0277-9536. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0277953623008031>.

GUPTA, A.; KUMAR, N.; BHATNAGAR, V. ANALYSIS OF MEDICAL DATA USING DATA MINING AND FORMAL CONCEPT ANALYSIS. Zenodo, nov. 2007. Disponível em: <https://doi.org/10.5281/zenodo.1073369>.

HU, J. X. et al. A large-cohort, longitudinal study determines precancer disease routes across different cancer types. Cancer Res., American Association for Cancer Research (AACR), v. 79, n. 4, p. 864–872, fev. 2019.

IGNATOV, D. et al. Triadic formal concept analysis and triclustering: searching for optimal patterns. Machine Learning - Springer, p. 271–302, 2015.

KARAMPATAKIS, V. et al. Comparison between normal values of 2-and 5-minute schirmer test without anesthesia. Cornea, LWW, v. 29, n. 5, p. 497–501, 2010.

KUMAR, C. A.; RADVANSKY, M.; ANNAPURNA, J. Analysis of a vector space model, latent semantic indexing and formal concept analysis for information retrieval. Cybernetics and Information Technologies, Versita, v. 12, n. 1, p. 34–48, 2012.

LANA, P. et al. Formal concept analysis applied to a longitudinal study of covid-19. In: INSTICC. Proceedings of the 24th International Conference on Enterprise Information Systems - Volume 2: ICEIS,. [S.l.]: SciTePress, 2022. p. 148–154. ISBN 978-989-758-569-2.

LEHMANN, F.; WILLE., R. A triadic approach to formal concept analysis. conceptual structures: ap-plications, implementation and theory. In: . [S.l.]: Springer, 1995.

LEMP, M. A.; FOULKS, G. N. The definition and classification of dry eye disease. Ocul Surf, v. 5, n. 2, p. 75–92, 2007.

LI, X. et al. Predictive effect of triglyceride glucose-related parameters, obesity indices, and lipid ratios for diabetes in a chinese population: A prospective cohort study. Front. Endocrinol. (Lausanne), Frontiers Media SA, v. 13, p. 862919, mar. 2022.

LORENZO, A. et al. The mental health impact of the covid-19 pandemic on post-secondary students: A longitudinal study. Psychiatry Research, v. 327, p. 115401, 2023. ISSN 0165-1781. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0165178123003517>.

MAIO, C. D. et al. Formal and relational concept analysis for fuzzy-based automatic semantic annotation. Applied Intelligence, v. 40, n. 1, p. 154–177, Jan 2014. ISSN 1573-7497.

MISSAOUI, R.; EMAMIRAD, K. Lattice miner-a formal concept analysis tool. In: 14th International Conference on Formal Concept Analysis. [S.l.: s.n.], 2017. p. 91.

MISSAOUI, R.; KUZNETSOV, S. O.; OBIEDKOV, S. A. (Ed.). Formal Concept Analysis of Social Networks. [S.l.]: Springer, 2017. (Lecture Notes in Social Networks). ISBN 978-3-319-64166-9.

MISSAOUI, R.; KWUIDA., L. Mining triadic associa-tion rules from ternary relations. In: Inter. Conf. on Formal Concept, Springer. [S.l.: s.n.], 2011. p. 204–218.

NEGM, E.; ABDELRAHMAN, S.; BAHGAT, R. PREFCA: A portal retrieval engine based on formal concept analysis. Information Processing & Management, Elsevier Ltd, v. 53, n. 1, p. 203–222, 2017. ISSN 03064573.

NORONHA, M. et al. Interpreting the human longevity profile through triadic rules - a case study based on the elsa-uk longitudinal study. In: STUD HEALTH TECHNOL INFORM. [S.l.]: PM, 2022. p. 782–786.

POELMANS, J.; ELZINGA, P.; DEDENE, G. Retrieval of criminal trajectories with an fca-based approach. In: NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS. PROCEEDINGS OF THE FCAIR 2013 FORMAL CONCEPT ANALYSIS MEETS INFORMATION RETRIEVAL WORKSHOP, CO-LOCATED WITH THE 35TH EUROPEAN CONFERENCE ON INFORMATION RETRIEVAL (ECIR 2013). [S.l.], 2013. v. 977, p. 83–94.

RIBEIRO, C. E. et al. A revision and analysis of the comprehensiveness of the main longitudinal studies of human aging for data mining research. WILEY INTERDISCIPLINARY REVIEWS: DATA MINING AND KNOWLEDGE DISCOVERY, Wiley Online Library, v. 7, n. 3, p. e1202, 2017.

RIBEIRO, C. R.; ZáRATE, L. E. Classifying longevity profiles through longitudinal data mining. EXPERT SYSTEMS WITH APPLICATIONS, v. 26, p. 75–89, 2019. ISSN 0957-4174.

ROTENSTEIN, L. S. et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: A systematic review and meta-analysis. JAMA, American Medical Association (AMA), v. 316, n. 21, p. 2214–2236, dez. 2016.

SELMANE, S. A. et al. Mining triadic association rules. COMPUTER SCIENCE & INFORMATION TECHNOLOGY, v. 3, p. 305–316, 2013.

SETIANINGRUM, F. et al. A prospective longitudinal study of chronic pulmonary aspergillosis in pulmonary tuberculosis in indonesia (APICAL). THORAX, BMJ, v. 77, n. 8, p. 821–828, ago. 2022.

SHAFER, R.; VUITTON, D. Highly active antiretroviral therapy (haart) for the treatment of infection with human immunodeficiency virus type 1. BIOMEDICINE PHARMACOTHERAPY, v. 53, n. 2, p. 73–86, 1999. ISSN 0753-3322. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0753332299800638>.

SHAMIR, R. et al. Analysis of blood-based gene expression in idiopathic parkinson disease. NEUROLOGY, AAN Enterprises, v. 89, n. 16, p. 1676–1683, 2017.

SINGH, P. K.; KUMAR, C. A.; GANI, A. A comprehensive survey on formal concept analysis, its research trends and applications. v. 26, n. 2, p. 495–516, 2016. ISSN 1641876X.

SINGH, P. K.; KUMAR, C. A.; GANI, A. A comprehensive survey on formal concept analysis, its research trends and applications. v. 26, n. 2, p. 495–516, 2016. ISSN 1641876X.

STUMME, G. et al. Computing iceberg concept lattices with titanic. DATA KNOWLEDGE ENGINEERING, p. 189–222, 2002.

STUMME, G.; WILLE, R.; WILLE, U. Conceptual knowledge discovery in databases using formal concept analysis methods. PROCEEDINGS OF THE SECOND EUROPEAN SYMPOSIUM ON PRINCIPLES OF DATA MINING AND KNOWLEDGE DISCOVERY, p. 450–458, 1998.

TRABELSI, C.; JELASSI, N.; YAHIA, S. B. Scalable mining of frequent tri-concepts from folksonomies. In: SPRINGER. Pacific-Asia Conference on Knowledge Discovery and Data Mining. [S.l.], 2012. p. 231–242.

WILKINSON, D. et al. Caloric vestibular stimulation for the management of motor and non-motor symptoms in parkinson's disease: Intention-to-treat data. Data in Brief, v. 25, p. 104228, 2019. ISSN 2352-3409. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352340919305827>.

WILLE, R. Why can concept lattices support knowledge discovery in databases? Proceedings of the concept lattices based knowledge discovery in databases workshop, p. 7–20, 2001.

WILLE, R. Restructuring lattice theory: an approach based on hierarchies of concepts. In: SPRINGER. International conference on formal concept analysis. [S.l.], 2009. p. 314–339.

YASSINE, S. et al. Functional brain dysconnectivity in parkinson's disease: A 5-year longitudinal study. Mov Disord, United States, v. 37, n. 7, p. 1444–1453, abr. 2022.

ZHENG, Z.; KOHAVI, R.; MASON, L. Real world performance of association rule algorithms. In: Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2001. (VLDB '94), p. 1–14. Disponível em: <https://doi.org/10.1145/502512.502572>.

ZHUK, R.; IGNATOV, D. I.; KONSTANTINOVA, N. Concept learning from triadic data. Procedia Computer Science, Elsevier, v. 31, p. 928–938, 2014.