Longitudinal Data Analysis Based on Triadic Rules to Describe of the Psychological Reactions During COVID 19 Pandemic

Lincoln A. N. Coutinho, Mark A. J. Song and Luis E. Zárate

Applied Computational Intelligence Laboratory – LICAP, Computer Science Department, Pontifical Catholic University of Minas Gerais, Brazil

Abstract: Longitudinal studies are essential to understand the evolution of individuals’ psychological behaviors, especially in pandemic scenarios. The work proposes the application of the triadic analysis, derived from the theory of Formal Analysis of Concepts, to describe, through rules, a longitudinal database about the attitudes and reactions of individuals during COVID 19. As a main result, one can observe how the different factors considered in the study are related in different scenarios of the pandemic, showing degrees of stress related to the prevention of the disease.

1 INTRODUCTION

A pandemic has various implications in a globalized society (Malta et al., 2020). Mental health is a highly relevant aspect, especially in an abnormal period where there may be a decrease in interpersonal contact due to imposed health restrictions. This aspect should not be downplayed, as it is correlated with public health issues such as anxiety, depression, distress, among other psychological problems. Therefore, studies on the mental health of a social group exposed to extreme situations, if used to identify behavioral patterns, can generate relevant information for public policies (Prati and Mancini, 2021). In this context, longitudinal studies can be employed to achieve these objectives. In this work, we present a method to identify psychological effects during the COVID-19 pandemic for the proposed longitudinal study in (O’Brien et al., 2021).

In general, longitudinal studies are used to investigate, for example, activities and behaviors of the same group of individuals over various periods of time, referred to as waves. Through periodic updating of records, it is possible to discover highly relevant patterns and temporal relationships from the available databases using data mining techniques and machine learning.

To analyze the longitudinal databases, we propose in this work to utilize the foundations of Formal Concept Analysis (FCA) theory (Ganter and Wille, 2012), a branch of applied mathematics based on the theory of conceptual lattices. Its main objective is to represent and extract knowledge from a dataset involving objects (individuals), attributes (clinical conditions, symptoms, etc.), and their incidence relations. This tuple of elements can be represented through a formal dyadic context, from which it is possible to extract association rules between the attributes. In summary, the main purpose of FCA is to summarize items in a database into information implications, such as the evidence of an individual symptom in a patient to better understand clinical diseases and their representation in society.

In general, FCA has been used in data analysis and knowledge representation, where associations and dependencies are identified from a binary incidence relationship between objects and attributes. Several works, such as (Carpineto and Romano, 2003), have discussed the use of extracting dyadic association rules through FCA.

Since datasets are often expressed by ternary and more generally n-ary relations, there has been a recent interest in proposing new solutions for the analysis and exploration of these multidimensional data, especially in triadic contexts (Bazin, 2020).

Triadic Concept Analysis (TCA), proposed by (Lehmann and Wille., 1995), is an extension of FCA theory that uses triadic formal context. This intro-
roduces a third element (to the dyadic context) - a condition determining the incidence relationship between objects and attributes. In the context of healthcare area, objects may correspond to patients, attributes to symptoms, and conditions to different waves of the longitudinal study. From the triadic context, it is possible to extract triadic association rules (Biedermann, 1999) that can be used to observe the relationships between attributes (symptoms) within a given wave or between waves of the longitudinal study. In this way, the TCA extends the FCA, showing not only the incidence of a symptom, but its recurrence in different scenarios (waves).

In (Trabelsi et al., 2012), the authors compared three algorithms for triadic data: TRICONS, TRIAS, and DATA-PEELER. In (Ignatov et al., 2015), the authors presented various strategies for discovering optimal patterns considering this type of data. In (Selmane et al., 2013), the authors proposed extracting triadic association rules by transforming the triadic context into an equivalent dyadic context. In (Zhuk et al., 2014), a comparison of algorithms for triadic context analysis is conducted, and the work by (Missouai and Emanirad, 2017) proposes a tool, Lattice-Miner, used in this work, for extracting triadic association rules.

The articles (Lana et al., 2022) and (Noronha et al., 2022) are likely the first works to explore triadic analysis in describing longitudinal study databases in the field of health. The first article deals with the analysis of the effectiveness of COVID-19 prevention procedures, and the second work focuses on pattern discovery related to human aging by observing the temporal evolution of clinical and environmental conditions of individuals. In this work, we propose the application of triadic rules to describe the psychological reactions of individuals to the COVID-19 pandemic based on a longitudinal study available in (O’Brien et al., 2021).

This article is divided into the following sections: in Section 2, the theoretical framework supporting the work is presented. In Section 3, the methodology used for discovering psychological patterns during the pandemic and the description of the dataset used are presented. Section 4 contains the experiments and analysis of the results from different scenarios. Finally, the conclusions and possible future work are outlined.

2 BACKGROUND

2.1 Formal Concept Analysis and Triadic Rules

Formal Concept Analysis (FCA) is a branch of applied mathematics related to the theory of conceptual lattices (Ganter and Wille, 2012). This theory is used to derive implicit relationships between objects and their attributes from a dataset. This relationship is formally defined by a tuple $K := (G, M, I)$, which is called a formal context, where $(I \subseteq G \times M)$, with $G$ representing objects, $M$ representing attributes, and $I$ representing incidence in the context—meaning which objects possess certain attributes, or, analogously, which attributes are present in certain objects. Within FCA, two main operators are defined:

$$A' := \{m \in M \mid \forall g \in A : (g, m) \in A\} \quad (1)$$

$$B' := \{g \in G \mid \forall m \in B : (g, m) \in B\} \quad (2)$$

Considering $A \subseteq G$ and $B \subseteq M$, a formal concept corresponds to the pair $(A, B)$, where $A$ represents a subset of objects (extension) and $B$ a subset of attributes (intention) in such a way that $A' = B$, and $B' = A$, where $A'$ and $B'$ are the derivation operators described by Equations 1 and 2.

Triadic Concept Analysis (TCA) (Wille, 1995) is defined by a quadruple $K := (G, M, I, C)$, where, in addition to objects and attributes, the incidence between them occurs under a condition $C$. For example, considering a longitudinal study with two waves, the start and end of a clinical treatment. Thus, a triadic concept represented by $(I \subseteq G \times M \times C)$ can be translated as objects that possess a certain subset of attributes under a condition, or similarly, attributes that are related to a subset of objects under a specific condition or wave.

From TCA, it is possible to generate two types of triadic implication rules known as BCAAR (Biedermann Conditional Attribute Association Rule) and BACAR (Biedermann Attributional Condition Association Rule), which have the following structure ((Biedermann, 1999)):

$$\text{BACAR: } (C_1 \rightarrow C_2) \cup M_1 [\text{support, confidence}]$$

$$\text{BCAAR: } (M_1 \rightarrow M_2) \cup C_1 [\text{support, confidence}]$$

The first rule, BACAR, occurs when $C_1, C_2 \subseteq C$ and $M_1 \subseteq M$. This means that when the subset of attributes $M_1$ occurs under condition $C_1$, it also occurs under condition $C_2$ with a certain support and confidence. The second rule, BCAAR, occurs when $M_1, M_2 \subseteq M$ and $C_1 \subseteq C$. This indicates that in the
wave $C_x$, the subset of attributes $M_1$ implies $M_2$ with a certain support and confidence. Our goal is to identify patterns of psychological relationships among individuals from the longitudinal database using both types of association rules.

3 METHODOLOGY

As mentioned, for this work, a longitudinal study aiming to assess the psychological conditions of individuals during the COVID-19 pandemic was considered. The first wave was conducted at the beginning of the pandemic between April 9th and 18th, 2020, and the second between June 19th and July 11th, 2020, approximately 2 months after the first data collection. In this database, several socioeconomic questions for individuals are present, highlighting aspects of isolation, medication, pre-existing conditions, psychological tests, working hours, among other issues (O’Brien et al., 2021).

The database comprises 151 interviewed individuals and 81 attributes, of which 12 were selected for a more detailed analysis of the results, totaling 24 for both waves. Tables I, II, and III describe the sets of questionnaires considered and the conditions in which they were conducted. It is worth noting that only the Five Facet Mindfulness Questionnaire (FFMQ) test, presented in Table I, has derivations, which is why Table II includes attributes related only to this specific scale test.

3.1 Methods

Preprocessing: For representing the longitudinal database in a triadic context, a discretization process is required. To achieve this, thresholds (reference values) had to be defined for marking the incidences of symptoms for each study wave. The objective is to describe the negative psychological influence of the pandemic, so values equal to or above the reference, which represent or are associated with a negative psychological condition, were considered and marked with 'X' for incidence, while values below this threshold did not determine an incidence and remained unmarked. A sample of this discretization is shown in Table IV.

For the PHQ15 (Patient Health Questionnaire Somatic Symptom Severity Scale), the literature-suggested scale was used as the reference threshold: Minimal (0-4), Low (5-9), Medium (10-14), and High (15-30). Specifically, the threshold value of 10, corresponding to the Medium level on the scale, was considered. For the IOES questionnaire, the study by (Weiss and Marmar, 1997) was used. The threshold was specifically set at 33, corresponding to the average value in the first wave of the study (O’Brien et al., 2021). For other questionnaires, thresholds were determined for extreme situations. To achieve this, an outlier analysis was performed using a box plot. This strategy aimed to define values deviating from the mean and variance that were associated with negative psychological conditions.

Subsequently, once the data were represented in the triadic concept, Lattice Miner ((Missaoui and Emamirad, 2017)), available on GitHub, was used for generating triadic rules. Before using the software, the database was converted to JSON (JavaScript Object Notation) format, containing all the relationships between objects, conditions, and attributes to be in the appropriate format for the software to generate the triadic rules. Consequently, the BCAARs and BACARs triadic rules can be generated based on desired confidence and support measures. The first, BCAARs can show the relation between two attributes (clinical conditions) in a certain wave (time) which could imply a certain dependence on two different symptoms. The second, BACARs, the recurrence of a attribute (clin-
Table 2: Derived Attributes Selected for Triadic Analysis.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFMQ Observe Mean</td>
<td>Weighted Mean of FFMQ Scale Items for Observation</td>
<td>Scale on how easy it is for the person to observe themselves in the last month</td>
</tr>
<tr>
<td>FFMQ Describe Mean</td>
<td>Weighted Mean of FFMQ Scale Questionnaire Items for Description</td>
<td>Scale on how easy it is for the person to describe themselves in the last month</td>
</tr>
<tr>
<td>FFMQ Aware Mean</td>
<td>Weighted Mean of FFMQ Scale Questionnaire Items for Attention</td>
<td>Scale on how easy it is for a person to focus on tasks in the last month</td>
</tr>
<tr>
<td>FFMQ Nonjudge Mean</td>
<td>Weighted Mean of FFMQ Scale Questionnaire Items for Judgment</td>
<td>Scale on the person’s self-judgment in the last month</td>
</tr>
<tr>
<td>FFMQ Nonreact Mean</td>
<td>Weighted Mean of FFMQ Scale Questionnaire Items for Reaction</td>
<td>Scale on the person’s reactions in the last month</td>
</tr>
</tbody>
</table>

Table 3: Conditions chosen for triadic analysis.

<table>
<thead>
<tr>
<th>Wave - Conditions</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (before)</td>
<td>April 9th to 18th, 2020</td>
<td>2 to 3 weeks after the beginning of the pandemics in the USA</td>
</tr>
<tr>
<td>D (after)</td>
<td>June 19th to July 11th, 2020</td>
<td>2 to 3 months after the condition “A” in the USA</td>
</tr>
</tbody>
</table>

Table 4: Transformation for Triadic Concept Analysis.

<table>
<thead>
<tr>
<th>Person ID</th>
<th>A - First Wave</th>
<th>D - Second Wave</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID 1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ID 2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ID 3</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ID 4</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

4.1 Scenario 1: Complete Database

In this context, 151 individuals were considered, and 1765 triadic rules were generated. This high number of rules is due to the fact that the rule explores all possible association rules within the dataset. In order to analyze the most significant rules, a minimum support of 30% was chosen for BACARs, and for BCAARs, a support of 25% was chosen:

**BACARS:**

\[ R_1 - (A \rightarrow D) \text{ IOES\_Total} \text{ [support = 36.9\% confidence = 80.9\%]} \]

\[ R_2 - (D \rightarrow A) \text{ IOES\_Total} \text{ [support = 36.9\% confidence = 88.7\%]} \]

\[ R_3 - (A \rightarrow D) \text{ PATS\_Mean} \text{ [support = 30.9\% confidence = 79.3\%]} \]

\[ R_4 - (D \rightarrow A) \text{ PATS\_Mean} \text{ [support = 30.9\% confidence = 68.7\%]} \]

**BCAARs:**

\[ R_1 - ( \text{ IOES\_Total} \rightarrow \text{ PATS\_Mean} ) A \text{ [support = 29.5\% confidence = 64.7\%]} \]

\[ R_2 - ( \text{ IOES\_Total} \rightarrow \text{ PATS\_Mean} ) D \text{ [support = 31.5\% confidence = 75.8\%]} \]

\[ R_3 - ( \text{ PATS\_Mean} \rightarrow \text{ IOES\_Total} ) A \text{ [support = 29.5\% confidence = 75.9\%]} \]

\[ R_4 - ( \text{ PATS\_Mean} \rightarrow \text{ IOES\_Total} ) D \text{ [support = 31.5\% confidence = 70.1\%]} \]

From the complete database, it was observed through BCAAR rules, \( R_1 \) and \( R_3 \), at the beginning of the pandemic, that people (29.5\%) who experienced strong stress symptoms in the last 7 days (IOES\_Total) also engaged in preventive actions against diseases (PATS\_Mean) with a confidence of 64.7\%. The reverse was also observed (PATS\_Mean \rightarrow IOES\_Total) in both study waves. This might suggest that concerning a pandemic, stress, which can be caused by traumatic events, may result in a person taking more preventive actions to avoid psychological conditions. Conversely, individuals accustomed to taking more preventive measures against diseases may also experience higher levels of stress.

4 EXPERIMENTS AND RESULTS

For the analysis, four different scenarios were chosen. The first scenario considers the complete database with employed and unemployed individuals. The remaining scenarios were subdivided into employed individuals and those not in the labor market. Employment conditions were selected based on the literature, including references such as (Borsoi, 2007) and (Hirschle and Gondim, 2020).

Each of the scenarios will be presented below, along with the most significant results observed in each of them.
4.2 Scenario 2: Individuals Employed Before the Pandemic

In this context, 138 individuals who were employed before the COVID-19 pandemic were analyzed. The minimum support threshold was set at 30%:

**BACARs:**

- **R1** - $(A \rightarrow D)$ IOES_Total [support = 40.4% confidence = 84.6%]
- **R2** - $(D \rightarrow A)$ IOES_Total [support = 40.4% confidence = 90.2%]
- **R3** - $(A \rightarrow D)$ PATS_Mean [support = 32.4% confidence = 80.0%]
- **R4** - $(D \rightarrow A)$ PATS_Mean [support = 32.4% confidence = 69.8%]

**BCAARs:**

- **R1** - $(IOES_Total \rightarrow PATS_Mean)$ A [support = 31.6% confidence = 66.2%]
- **R2** - $(IOES_Total \rightarrow PATS_Mean)$ D [support = 33.8% confidence = 75.4%]
- **R3** - $(PATS_Mean \rightarrow IOES_Total)$ A [support = 33.6% confidence = 78.2%]
- **R4** - $(PATS_Mean \rightarrow IOES_Total)$ D [support = 33.6% confidence = 71.2%]

Analyzing the results for individuals who were employed before the pandemic, the outcome was very similar to the complete database (section 4.1). The only difference is that the support and confidence levels are generally higher, meaning the relationship between stress and disease prevention worsened even further in this scenario.

4.3 Scenario 3: Individuals who Were Working at the Beginning of the Pandemic

In this scenario, 127 individuals who were working at the beginning of the COVID-19 pandemic (April 9th to 18th, 2020) were analyzed. BACARs and BCAARs were analyzed with a minimum support of 30%:

**BACARs:**

- **R1** - $(A \rightarrow D)$ IOES_Total [support = 41.6% confidence = 83.9%]
- **R2** - $(D \rightarrow A)$ IOES_Total [support = 41.6% confidence = 92.9%]

**BCAARs:**

- **R1** - $(IOES_Total \rightarrow PATS_Mean)$ A [support = 33.6% confidence = 67.7%]
- **R2** - $(IOES_Total \rightarrow PATS_Mean)$ D [support = 33.6% confidence = 75.0%]
- **R3** - $(PATS_Mean \rightarrow IOES_Total)$ A [support = 33.6% confidence = 80.8%]
- **R4** - $(PATS_Mean \rightarrow IOES_Total)$ D [support = 33.6% confidence = 71.2%]

Analyzing the results for individuals who were working at the beginning of the pandemic, the outcome was very similar to the complete database (section 4.1) and to individuals who were employed before the pandemic (section 4.2). However, the support and confidence levels are mostly even higher, implying a deterioration in terms of stress, disease prevention, and the relationship between these two factors.

4.4 Scenario 4: Individuals who were Not Employed at the Beginning of the Pandemic

In this scenario, 24 individuals who were not employed at the beginning of the COVID-19 pandemic (April 9th to 18th, 2020) were analyzed, and the BACARs and BCAARs had a minimum support of 25%:

**BACARs:**

- **R1** - $(A \rightarrow D)$ FFMQ_NonreactMean [support = 25.0% confidence = 66.7%]
- **R2** - $(D \rightarrow A)$ FFMQ_NonreactMean [support = 25.0% confidence = 50.0%]
- **R3** - $(A \rightarrow D)$ PVD_Total [support = 29.2% confidence = 100.0%]
- **R4** - $(D \rightarrow A)$ PVD_Total [support = 29.2% confidence = 87.5%]

**BCAARs:**

- **R1** - $(FFMQ_NonreactMean \rightarrow PATS_Mean)$ A [support = 25.0% confidence = 50.0%]
- **R2** - $(PATS_Mean \rightarrow FFMQ_NonreactMean)$ A [support = 25.0% confidence = 75.0%]
Considering the BCAAR rules, R1 and R2, for individuals who were not employed at the beginning of the pandemic, the rules associate preventive actions that individuals can take against COVID-19 (PATS_Mean) with attention-related factors such as non-reactivity (FFMQ_NonreactMean), where non-reactivity to inner experience is defined in terms of allowing thoughts and feelings to come and go without being caught up or carried away by them, where a higher score indicates higher attention (Bohlmeijer et al., 2011). It can be observed that all individuals with 75% preventive actions were non-reactive. In this scenario, it was only highlighted that the major issues for individuals who were not employed at the beginning of the pandemic are different compared to employed individuals analyzed in the previous scenarios.

5 CONCLUSIONS AND FUTURE WORK

The objective of this study was to demonstrate the potential of triadic analysis for extracting association rules within the context of longitudinal studies for psychological records, focusing on people’s reactions to stress conditions such as the COVID-19 pandemic. The rules can contribute to a better understanding of individuals’ psychological reactions under stressful conditions.

It is important to emphasize that the approach requires prior definition of thresholds for discretization and determining whether an incidence is marked or not in the triadic context. The difficulty in accessing information about the tests on the applied scale, due to restricted data, can hinder threshold definition to characterize individuals as healthy or not, and to better understand the topic being addressed. Although threshold definition requires expert knowledge, the approach allows for the adjustment of various scenarios to describe the results of a longitudinal study. Among the positive aspects, the applicability and ease of application to various contexts can be highlighted.

As future work, there are several implementations that can be carried out in the considered longitudinal study. Implementations can be performed in different scenarios, and different attributes from the database can be used in each of these chosen scenarios. Furthermore, with the input of researchers from the field of psychology, the data could be better analyzed to understand the implications of a pandemic on people’s mental health. Our intention was to demonstrate the potential use of triadic rules to analyze the psychological effects of a pandemic.

ACKNOWLEDGEMENTS

The authors would like to thank: The National Council for Scientific and Technological Development of Brazil (CNPQ), The Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) (Grant PROAP 88887.842889/2023-00 – PUC/MG, Grant PDPG 88887.708960/2022-00 – PUC/MG - INFORMATICA and Finance Code 001), Minas Gerais State Research Support Foundation (FAPEMIG) under grant number APQ-01929-22, and the Pontifical Catholic University of Minas Gerais, Brazil.

REFERENCES


