

Article

Bridging Crisp-Set Qualitative Comparative Analysis and Association Rule Mining: A Formal and Computational Integration

Acácio Dom Luís , Rafael Benítez *  and María del Carmen Bas 

Departamento de Matemáticas para la Economía y la Empresa, Facultad de Economía, Universidad de Valencia, Avda. Tarongers s/n, 46022 Valencia, Spain; adomluis@alumni.uv.es (A.D.L.); maria.c.bas@uv.es (M.d.C.B.)

* Correspondence: rafael.suarez@uv.es

Abstract: In this paper, a novel mathematical formalization of Crisp-Set Qualitative Comparative Analysis (csQCA) that enables a rigorous connection with a specific class of association rule mining (ARM) problems is proposed. Although these two methodologies are frequently used to identify logical patterns in binary datasets, they originate from different traditions. While csQCA is rooted in set theory and Boolean logic and is primarily applied in the social sciences to model causal complexity, ARM originates from data mining and is widely used to discover frequent co-occurrences among items. In this study, we establish a formal mathematical equivalence between csQCA configurations and a subclass of association rules, including both positive and negative conditions. Moreover, we propose a minimization procedure for association rules that mirrors the Quine–McCluskey reduction method employed in csQCA. We demonstrate the consistency of the results obtained using both methodologies through two examples (a small-N study on internet shutdowns in Sub-Saharan Africa and a large-N analysis of immigration attitudes in Europe) and some numerical experiments. However, it is also clear that ARM offers improved scalability and robustness in high-dimensional contexts. Overall, these findings provide researchers with valuable theoretical and practical guidance when choosing between these approaches in qualitative data analysis.

Keywords: qualitative comparative analysis; association rule mining; Boolean logic; causal inference

MSC: 68Txx; 91Fxx; 00A06



Academic Editor: Raymond Lee

Received: 2 May 2025

Revised: 29 May 2025

Accepted: 9 June 2025

Published: 11 June 2025

Citation: Dom Luís, A.; Benítez, R.; Bas, M.d.C. Bridging Crisp-Set Qualitative Comparative Analysis and Association Rule Mining: A Formal and Computational Integration. *Mathematics* **2025**, *13*, 1939. <https://doi.org/10.3390/math13121939>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

While Qualitative Comparative Analysis (QCA) and association rule mining (ARM) are both widely used techniques for uncovering logical patterns in general datasets, they originate from distinct research traditions. QCA, especially in its Crisp-Set form (csQCA), was developed by Charles Ragin in the 1980s [1] as a set-theoretic method for analyzing causal complexity in small- to medium-sized case studies, particularly in the social sciences. It enables researchers to identify configurations of causal conditions that are necessary or sufficient for a given outcome, emphasizing multiple factors' interplay and conjunctural nature [2]. This method is particularly valuable for its ability to handle causal complexity and its focus on identifying patterns between cases rather than isolating individual variables [3]. A standard solution configuration in a QCA problem is characterized by a causal relationship of the form “If this combination of factors is given, then this outcome occurs”.

In contrast, ARM emerged from the field of data mining and was initially used in market basket analysis to discover frequent item co-occurrences in transactional databases [4].

The technique identifies rules of the form “if items A and B occur, then item C also occurs,” using metrics such as support and confidence to evaluate the strength and reliability of such patterns. Over the years, ARM has been applied in diverse fields, including marketing, bioinformatics, healthcare, and finance, due to its scalability and computational efficiency.

Despite the conceptual similarities between csQCA and ARM—with both generating if–then rules from binary data using Boolean logic—their methodological foundations and evaluation strategies differ substantially. Surprisingly, few studies have formally examined their theoretical relationship, and even fewer have attempted to unify their respective solution processes. This lack of integration represents a gap in the literature, especially as researchers increasingly turn to hybrid or comparative methodological approaches.

This study aims to bridge this gap by establishing a formal equivalence between csQCA configurations and a subset of association rules that include both positive and negative conditions. We show that any csQCA problem can be reformulated as an ARM problem under a specific encoding of presence and absence conditions. Moreover, we prove that the key evaluation metrics in csQCA—consistency and coverage—are mathematically equivalent to the support and confidence measures used in ARM.

The contributions of this paper are threefold. First, we propose a mathematical framework for csQCA that facilitates precise mapping to association rule structures. Second, we introduce a theorem that enables the minimization of association rules similarly to the Quine–McCluskey logic minimization procedure employed in csQCA, enabling the identification of reduced rule sets without relying on exponentially large truth tables. Third, we illustrate our theoretical findings through two comparative applications: a small-N study on internet shutdowns during African elections and a large-N analysis of opposition to immigration in Europe.

Through formally connecting csQCA and ARM, we provide researchers with a deeper understanding of the conditions under which these methods yield similar or complementary insights. Our findings also demonstrate the advantages of ARM in terms of robustness and scalability, particularly in high-dimensional datasets. This opens up new opportunities for researchers to integrate or select between these approaches more effectively, depending on the nature and size of their empirical data.

The remainder of this paper is structured as follows: In Section 2, we provide a brief review of the relevant literature on csQCA and association rule mining (ARM). Section 3 presents our main theoretical contributions: we introduce a new mathematical formalization of csQCA (Section 3.1), outline the ARM framework (Section 3.2), and prove an equivalence theorem linking csQCA configurations to a specific class of association rules (Section 3.3). We also propose an ARM minimization procedure analogous to the Quine–McCluskey reduction used in csQCA (Section 3.4). Section 4 illustrates these ideas through two empirical applications (a small-N study on internet shutdowns and a large-N analysis of immigration attitudes) and a numerical computational experiment. Section 5 discusses the theoretical and practical implications of the results. Finally, in Section 6, we provide some conclusions and directions for future research.

2. Literature Review

In recent times, researchers from different areas of knowledge have opted for the use of configurational comparative methods. Configurational methods propose different types of solutions and new ways of approaching problems [5,6]. The increasing adoption of these methodologies by researchers is justified by the fact that they provide powerful methodological tools to find frequent patterns in datasets. Therefore, some of these tools focus on inferring cases as condition configurations and using truth tables [7], such as the Qualitative Comparative Analysis (QCA) technique, and although the data configuration in

the association rule mining (ARM) algorithms presents a truth table structure, the solution search process is based on solution metrics of interest [4] (i.e., a solution may be prevalent but not interesting). Similarly, there have been doubts about the best solution for each type of research problem [8].

The Crisp-Set QCA (csQCA) methodology is used in social research to analyze and identify combinations of causal conditions that lead to a specific outcome. It combines qualitative and quantitative techniques to examine complex patterns and identify causal configurations that can explain social phenomena [9]. Meanwhile, the ARM methodology is widely used to discover interesting patterns in datasets. These patterns are expressed as association rules that indicate the relationship between different elements in a set of transactions [4].

In the last 30 years, numerous papers have been published in different areas of knowledge using both methodologies [10–17]. These methodologies have accentuated applications in several fields, such as engineering, mathematics, medicine, computer science, administration and management, political science, economics, finance, biology, and education [14,18].

Many authors have advocated for the use of configurational methods instead of classical statistical methods [19–21]. These researchers argue that configurational theory is fundamentally grounded in the assumption that combinations of initial variables can lead to the same outcome. Consequently, the relationship between an outcome and its preconditions is often more asymmetric than symmetric. For instance, different groups of customers may choose to fly with a particular airline, considering different variables or attributes. However, certain customers may decide not to fly with an airline until a specific condition is met, even if the provided condition alone does not result in the intended use. Such asymmetric relationships and combinatorial complexities cannot be modeled using conventional regression-based methods.

According to [22], a configuration is a specific set of causal variables that interact, producing a desired outcome. However, using data analysis techniques that focus on causal configurations helped estimate the causal contribution of various configurations/combinations to achieve the expected outcome.

Table 1 presents studies by various authors that aim to compare or integrate different methodologies, elucidating the parallels, distinctions, strengths, and weaknesses of classical statistical techniques compared to emerging configurational methods. Notable examples within these studies include the examination of QCA versus linear regression models (LRMs) [19,23,24], AR versus LRMs [25,26], QCA versus Structural Equation Models (SEMs) [27], and AR versus SEMs [28]. This table furnishes a concise summary of the most significant investigations comparing configurational methods and classical statistical approaches.

In recent years, there has been an effort to integrate QCA with computational or rule-based approaches, and the following studies were identified through an extensive review of the literature. One study proposed a hybrid model that combines QCA with Support Vector Machines (SVMs), utilizing various kernels and ensemble techniques (Bagging, Random Subspace, and AdaBoost) to analyze passenger resistance to biometric e-gates at airports [29]. This approach integrated QCA's configurational logic with the predictive capabilities of machine learning algorithms. A further study introduced a Bayesian rule-based model as a quantitative alternative to QCA, capturing complex causal relationships through probabilistic logic [30]. The integration of QCA with process tracking has been explored to strengthen causal inference in case-based research. In this approach, conservative QCA solutions provide a foundation for more in-depth research [31]. Simultaneously, the development of the *scpQCA* method is proposed [32]. The *scpQCA* method constitutes an

extension of multi-value QCA (mvQCA) grounded in set-covering principles. The objective of the method's development *scpQCA* is to overcome the typical limitations of mvQCA, such as low coverage and limited interpretability. More recently, a 2025 investigation applied QCA within an explanatory sequential mixed-methods design, integrating quantitative and qualitative analyses to examine family-related factors influencing the outcomes of parenting programs [33].

Table 1. An overview of selected studies comparing or combining QCA, association rules (ARs), Logistic Regression Models (LRM), and Structural Equation Modeling (SEM).

Ref	Area	Sample Size	Objectives	QCA	AR	LRM	SEM	Main Findings
Ragin et al. (2003) [19]	Sociology	41 villages in Southern India	Analyze the strengths and weaknesses of QCA and compare with logistic regression.	✓		✓		QCA emphasizes combinatorial complexity; stats focus on net effects. QCA suits moderate <i>n</i> and is error-sensitive. Both can complement each other.
Seawright (2005) [24]	Political Science	~80–100 electoral processes in Latin America (1980–1996)	Compare QCA with regression analysis regarding assumptions for causal inference.	✓		✓		QCA requires assumptions at least as stringent as those used in regression analysis.
Buxton et al. (2019) [25]	Biomedicine	65.4K patients (2013–2016)	Discover unknown correlations using AR and logistic regression.		✓	✓		AR identifies patterns; LRM removes false associations, enhancing clinical relevance.
Changpetch (2013) [26]	Computer Science	432 robots	Model selection through optimal combinations of main effects and interactions using AR.		✓	✓		Combining both methods improves model fit and interpretability.
Zhao and Yan (2020) [27]	Computer Science	229 Internet users	Combine SEM and QCA to analyze individual and combined variable effects.	✓			✓	Their combination yields more compelling result interpretations.
Vis (2012) [23]	Political Science	53 governments in 18 democracies (1985–2003)	Assess pros and cons of QCA and regression models.	✓		✓		Both have strengths and drawbacks for moderate- <i>n</i> studies.

In addition to using Crisp-Set QCA, other variants, such as fsQCA and mvQCA, have been used in various domains. This case study employed fsQCA to identify configurations that contribute to adherence to development timelines in Product Lifecycle Management (PLM) co-development projects [34]. In a similar study, fsQCA was applied to assess stakeholder satisfaction in public participation processes under the EU Water Framework Directive [35]. In corporate governance, fsQCA was used to examine the interactive effects of good governance practices and CEO profiles on ESG performance [36]. These studies underline the adaptability of fsQCA in managing intricate causal relationships across diverse contexts.

Notwithstanding these advances, the existing literature contains no study formally establishing mathematical equivalence between csQCA and ARM. To the best of our knowledge, there has been no prior attempt to implement a direct one-to-one transformation between csQCA configurational solutions and association rules encompassing both presence and absence conditions or to provide formal proofs of equivalence between key evaluation metrics, including *consistency* and *coverage* in QCA and *support* and *confidence* in ARM. This study is the first systematic attempt to formally integrate csQCA and ARM, thereby providing a coherent theoretical model and substantial computational advantages, including scalability, interpretability, and applicability to high-dimensional contexts. These advantages are supported by empirical applications involving both small-*N* and large-*N* datasets.

3. Materials and Methods

3.1. QCA

QCA is a prominent methodology based on solid foundations in set theory. The first version [1] presented a binary system called Crisp Sets (csQCA), which was eventually extended to Fuzzy Sets (fsQCA) and multivalued variants, known as mvQCA [8].

QCA conceptualizes causal relationships between social phenomena as established relationships. This perspective enables the analysis of causal complexity by constructing and examining arguments about the necessity and/or sufficiency of causal conditions—combinations of attributes theoretically relevant to the study—for a given causal outcome [37].

In this section, we present a novel mathematical formalization of the Crisp-Set Qualitative Comparative Analysis (csQCA) methodology. While csQCA is typically described in set-theoretic and logical terms, our objective here is to provide a precise formal framework that will allow us, in the following sections, to establish a rigorous correspondence between csQCA and a specific class of association rule mining (ARM) problems.

This formalization will serve as the foundational structure for proving a theoretical equivalence between csQCA solution configurations and association rules, including both positive and negative conditions.

Definition 1. A csQCA model is a tern $(\mathcal{V}, R, M)$, where $\mathcal{V} = \{V_1, \dots, V_n\}$ is a set of n variables; R is an outcome or response variable (i.e., another variable, such that $R \notin \mathcal{V}$); and M is the sample dataset, which is an $m \times (n + 1)$ matrix of the form

$$M = \begin{array}{c|cccc|c} & V_1 & V_2 & \cdots & V_n & R \\ \hline \mathbf{c}_1 & x_{11} & x_{12} & \cdots & x_{1n} & y_1 \\ \mathbf{c}_2 & x_{21} & x_{22} & \cdots & x_{2n} & y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{c}_m & x_{m1} & x_{m2} & \cdots & x_{mn} & y_m \end{array}, \quad (1)$$

where $x_{ij}, y_i \in \{0, 1\}$, for all $i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$, where 1 indicates the presence and 0 indicates the absence of each condition/variable and the outcome.

Each row of the matrix M represents a sample **case** and is denoted by

$$\mathbf{c}_i = (\mathbf{x}_i; y_i), \quad i = 1, \dots, m,$$

where $\mathbf{x}_i = (x_{i1}, \dots, x_{in})$ is the antecedent vector, and y_i is the consequent (or outcome) value.

Definition 2. We say that a csQCA dataset M is **consistent** if, for any given two cases, $\mathbf{c}_{i_1} = (\mathbf{x}_{i_1}; y_{i_1})$ and $\mathbf{c}_{i_2} = (\mathbf{x}_{i_2}; y_{i_2})$, if $\mathbf{x}_{i_1} = \mathbf{x}_{i_2}$; then, necessarily, $y_{i_1} = y_{i_2}$. This means that there are no contradictions among the sample cases.

Definition 3. A **configuration** of order k ($k \leq n$) is a pair $\mathcal{S}_k = (\mathcal{V}_k, \mu)$, where $\mathcal{V}_k \subset \mathcal{V}$ with $|\mathcal{V}_k| = k$ and $\mu : \mathcal{V}_k \rightarrow \{0, 1\}$.

Definition 4. A case $\mathbf{c}_i = (\mathbf{x}_i; y_i)$ is **compatible** with the configuration $\mathcal{S}_k = (\mathcal{V}_k, \mu)$ if

$$\mu(V_{j_l}) = x_{ij_l}, \quad \forall l \in \{1, \dots, k\}$$

Definition 5. A configuration $\mathcal{S}_k$ is **conducive** to the outcome R if there exists at least a case $\mathbf{c}_i = (\mathbf{x}_i; y_i)$ compatible with $\mathcal{S}_k$ for which $y_i = 1$.

Remark 1. It is important to note that all possible n -configurations (conductive or non-conductive to the outcome R) can be written as a truth table with 2^{n+1} rows. Each row of the truth table corresponds to a specific combination of conditions, and collectively, all the rows display the possible logical combinations of conditions using the logical AND operation. These configurations are observed in one or more cases (the rows of the matrix M), and sufficiency is determined by the association of these cases with the outcome R [38].

An important aspect of QCA analysis is the reducibility or minimization of configurations. This is particularly essential because it enables the simplification of complex configurations into more parsimonious and interpretable expressions without losing their explanatory power. This process relies on a fundamental property of set theory $(A \cap B) \cup (A \cup \sim B) = A$. In other words, when a condition A leads to an outcome both in the presence (B) and in the absence ($\sim B$) of another condition, the influence of B becomes irrelevant and can be eliminated. Thus, minimization identifies and removes such redundancies across cases, producing a simpler and more general causal model that still accurately captures the empirical patterns observed. Next, we define the concept of minimizable configurations within the framework of our mathematical formalization.

Definition 6. Two configurations of order k , $\mathcal{S}_k^{(1)} = (\mathcal{V}_k^{(1)}, \mu^{(1)})$ and $\mathcal{S}_k^{(2)} = (\mathcal{V}_k^{(2)}, \mu^{(2)})$, can be **minimized** (or **reduced**) to a configuration of order $k - 1$ if

- Both are conducive to the outcome R ;
- $\mathcal{V}_k^{(1)} = \mathcal{V}_k^{(2)} = \{V_{j_1}, \dots, V_{j_k}\}$;
- There exists a $p \in \{1, \dots, k\}$ such that

$$|(\mu_1(V_{j_1}), \mu_1(V_{j_2}), \dots, \mu_1(V_{j_k})) - (\mu_2(V_{j_1}), \mu_2(V_{j_2}), \dots, \mu_2(V_{j_k}))| = \mathbf{e}_p, \quad (2)$$

where $\mathbf{e}_p \in \mathbb{R}^k$ is the p -th vector of the canonical base (i.e., $\mathbf{e}_p = (0, \dots, 0, 1, 0, \dots, 0)$).

In that case, $\mathcal{S}_k^{(1)}$ and $\mathcal{S}_k^{(2)}$ can be reduced to the $(k - 1)$ -configuration $\mathcal{S}_{k-1} = (\mathcal{V}_{k-1}, \mu)$, where

$$\mathcal{V}_{k-1} = \mathcal{V}_k^{(1)} \setminus \{V_{j_p}\} = \{V_{j_1}, \dots, V_{j_{p-1}}, V_{j_{p+1}}, \dots, V_{j_k}\},$$

and

$$\mu = \mu^{(1)}|_{\mathcal{V}_{k-1}} = \mu^{(2)}|_{\mathcal{V}_{k-1}}.$$

It is important to note that from (2), it is inferred that for every $V \in \mathcal{V}_{k-1}$, $\mu^{(1)}(V) = \mu^{(2)}(V)$.

If two configurations cannot be minimized, we say they are **irreducible**.

The objective in csQCA is to obtain all irreducible configurations leading to a certain outcome. That is, we can define the solution set of a csQCA problem as follows:

Definition 7. The **solutions** of the csQCA model $(\mathcal{V}, R, M)$ are given by

$$\mathcal{C} = \bigcup_{k=1}^n \mathbf{C}_k$$

where each set $\mathbf{C}_k$ is the family of all irreducible configurations of order k ($\mathbf{C}_k$ may be void) conducive to the outcome R ; that is, $\mathbf{C}_k$ is given by

$$\mathbf{C}_k = \bigcup_{p=1}^{n_k} \mathcal{S}_k^{(p)},$$

where

- $\mathcal{S}_k^{(p)}$ is conducive to outcome R for all $p \in \{1, \dots, n_k\}$.
- $\mathcal{S}_k^{(p_1)}$ and $\mathcal{S}_k^{(p_2)}$ are irreducible $\forall p_1, p_2 \in \{1, \dots, n_k\}$ with $p_1 \neq p_2$.

In csQCA, evaluating the reliability and relevance of causal configurations requires the use of classical metrics, namely, **consistency** and **coverage**. These metrics help to determine the strength and explanatory power of the identified patterns. Consistency measures the degree to which cases supporting a given configuration also exhibit the outcome, ensuring that the solution is logically coherent with the dataset. That is,

Definition 8. Given a configuration $\mathcal{S}_k = (\mathcal{V}_k, \mu)$, with $\mathcal{V}_k = \{V_{j_1}, V_{j_2}, \dots, V_{j_k}\}$, then the **consistency** of the configuration is defined as

$$\text{cons}(\mathcal{S}_k) = \frac{n_{xy}}{n_x}, \quad (3)$$

where

$$\begin{aligned} n_x &= |\{\mathbf{c} = (\mathbf{x}, y) : \mathbf{c} \text{ is compatible with } \mathcal{S}_k\}| \\ n_{xy} &= |\{\mathbf{c} = (\mathbf{x}, y) : \mathbf{c} \text{ is compatible with } \mathcal{S}_k \text{ and } y = 1\}| \end{aligned} \quad (4)$$

That is, n_x represents the number of cases in the sample dataset compatible with the configuration $\mathcal{S}_k$, regardless of the outcome of the case; on the other hand, n_{xy} is the number of cases in the sample dataset compatible with $\mathcal{S}_k$ for which the outcome is positive.

However, coverage assesses how much of the observed outcome is explained by a specific configuration, providing insight into the generalizability of the findings, as described in the following definition.

Definition 9. The **coverage** of a causal configuration is defined as the proportion of cases with a positive outcome compatible with the given configuration. That is,

$$\text{cov}(\mathcal{S}_k) = \frac{n_{xy}}{n_y}, \quad (5)$$

where n_{xy} is defined above, and n_y is the number of cases in the sample dataset for which the result is 1, i.e.,

$$n_y = |\{\mathbf{c} = (\mathbf{x}, y) : y = 1\}| \quad (6)$$

A high-consistency configuration indicates a strong causal link, while high coverage suggests that the configuration accounts for a significant portion of the cases with the outcome.

A configuration $\mathcal{S}_k$ is compatible with the sample dataset M if it is compatible with at least one case of the sample dataset. Otherwise, we say that the configuration and the sample dataset are incompatible.

We obviously have that, given a configuration $\mathcal{S}_k$,

1. If $\mathcal{S}_k$ is compatible with M , then $\text{cons}(\mathcal{S}_k) > 0$.
2. If $\mathcal{S}_k$ is conducive to the outcome R , and the sample dataset M is consistent (there are no contradictions), then $\text{cons}(\mathcal{S}_k) = 1$.

The procedure for finding the solution set $\mathcal{C}$ is based on the Boolean minimization algorithm developed in the 1950s by Quiney and McCluskey in the context of electronic circuit optimization [39–41]. This procedure is a top–bottom algorithm that starts with the maximal configurations—that is, the configurations of order n (where n is the number

of variables in the set $\mathcal{V}$)—and then, for each size k , it finds the configurations of order k reducible to configurations of order $k - 1$. The procedure stops once the configurations of order 1 are reached, or at a certain step, the configurations of order k are the empty set.

The computation of csQCA solutions relies on analyzing truth tables, which systematically enumerate all possible configurations of causal conditions. Although this approach ensures a comprehensive evaluation of causal relationships, it also presents a significant computational challenge. As the number of possible configurations grows exponentially with the number of variables (specifically, as 2^n for n binary variables), analyzing large datasets with numerous causal conditions becomes increasingly infeasible. This combinatorial explosion can lead to excessive processing times and computational limitations, especially in cases where the dataset includes dozens of variables.

3.2. Association Rules Mining Model

Association rule mining (ARM), formalized by Agrawal et al. (1993) [4], is a data-driven methodology designed to identify frequent item sets and extract meaningful if–then rules from transactional datasets. These rules, often expressed as $X \Rightarrow Y$, reveal co-occurrence patterns where the presence of item set X implies the presence of item set Y , quantified by metrics such as **support** (the frequency of co-occurrence) and **confidence** (conditional probability). Originally developed for market basket analysis, ARM has evolved into a cornerstone technique for exploratory data mining in diverse domains, including retail, healthcare, and bioinformatics [18,42].

The widespread adoption of ARM in recent decades stems from its algorithmic efficiency in handling large-scale datasets. Central to this is the Apriori algorithm [43], which employs a breadth-first search strategy to iteratively generate candidate item sets and prune those failing to meet user-defined minimum support thresholds. By leveraging the downward closure property (subsets of frequent item sets are themselves frequent), Apriori reduces computational complexity, enabling scalable pattern discovery even in high-dimensional datasets [18].

Mathematically, ARM operates on a transactional database T , where each transaction is represented by a binary vector encoding the presence (1) or absence (0) of items from a universal set I . More precisely, let $I = \{I_1, I_2, \dots, I_m\}$ be a universal set of **items**, and let $T = \{\tau_1, \tau_2, \dots, \tau_n\}$ be a transactional database where each transaction $\tau_i \subseteq I$. Then, we have the following:

Definition 10. An association rule [4] is an implication of the form

$$X \Rightarrow Y \quad \text{where} \quad X \subseteq I, Y \subseteq I, X \cap Y = \emptyset, \text{ and } X, Y \neq \emptyset.$$

This rule indicates that the occurrence of item set X in a transaction implies the occurrence of item set Y in the same transaction; that is, for every transaction $\tau \in T$,

$$X \subseteq \tau \longrightarrow Y \subseteq \tau.$$

Analogously to csQCA problems, extracting meaningful association rules relies on quantitative criteria to evaluate their statistical significance and practical relevance. Two fundamental metrics serve as critical filters to distinguish robust patterns from spurious correlations: **support** and **confidence**. Support ensures that rules reflect frequently observed co-occurrences, while confidence measures the reliability of the implication. Mathematically,

Definition 11. Given an association rule $X \Rightarrow Y$, we define

- **Support:** The proportion of transactions in T containing both X and Y :

$$\text{supp}(X \Rightarrow Y) = \frac{|\{\tau \in T \mid X \cup Y \subseteq \tau\}|}{|T|}.$$

- **Confidence:** The conditional probability that a transaction containing X also contains Y :

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)} = \frac{|\{\tau \in T \mid X \cup Y \subseteq \tau\}|}{|\{\tau \in T \mid X \subseteq \tau\}|}.$$

These metrics enable researchers to prioritize rules that are both prevalent in the dataset and probabilistically robust, thereby balancing generality and predictive strength. By setting thresholds for these values, practitioners can systematically discard trivial or coincidental associations, focusing instead on actionable insights with empirical validity.

In an association rule $X \Rightarrow Y$, X is the **antecedent** (left-hand side, LHS), and Y is the **consequent** (right-hand side, RHS).

Definition 12. The rule $X \Rightarrow Y$ is considered **strong** or **interesting** if it satisfies user-defined thresholds for both support (frequency) and confidence (reliability), commonly referred to as minimum support and minimum confidence.

In practice, identifying interesting rules (those satisfying minimum support and confidence thresholds) relies on efficient algorithms capable of handling large datasets. One of the most widely used methods for this purpose is the Apriori algorithm, which systematically uncovers frequent item sets that serve as the foundation for generating such rules.

The Apriori algorithm is a foundational technique in association rule mining (ARM) used to discover frequent item sets and generate association rules from large datasets. It operates by systematically identifying item sets whose frequency exceeds a predefined minimum support threshold. From these frequent item sets, association rules are generated, provided their confidence surpasses a user-specified minimum level—thus ensuring both frequency and reliability.

The algorithm proceeds iteratively: it begins with 1 item set and expands to larger item sets by joining previously discovered frequent item sets. To maintain computational efficiency, Apriori leverages the **downward closure property**, which states that if an item set is frequent, then all of its subsets must also be frequent. Conversely, if an item set is infrequent, all of its supersets can be safely pruned from further consideration.

Despite its effectiveness, the Apriori algorithm can become computationally expensive with large datasets due to the generation of numerous candidate item sets. To address this, several alternative algorithms have been developed to enhance efficiency and scalability. Notable alternatives include the following:

- FP-Growth (Frequent Pattern Growth), which eliminates the need for candidate generation by using a compact tree structure to store transactions [44].
- ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal), which relies on a depth-first search strategy and intersection operations to find frequent item sets [45].
- CHARM (Closed Association Rule Mining), which focuses on discovering closed item sets to reduce redundancy [46].
- RARM (Rapid Association Rule Mining), optimized for high-dimensional datasets by leveraging specialized data structures [47].

These advanced algorithms offer performance improvements over Apriori, particularly in handling large-scale and high-dimensional data scenarios.

3.3. The Equivalence Theorem

Having established the mathematical formalization of the two techniques—csQCA and AR—we now show that any csQCA problem can be reformulated as a certain type of association rule mining problem. In particular, we use the so-called negative association rules, which capture relationships involving the absence of an item [48,49].

Let us consider a csQCA model $(\mathcal{V}, R, M)$, with $\mathcal{V} = \{V_1, \dots, V_n\}$ and $M = (x_{ij}; y_i)$, for $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$. Then, we define the following AR problem:

- The item set is $I = \{V_1, v_1, V_2, v_2, \dots, V_n, v_n, R, r\}$, with $2n + 2$ items.
- The transaction database is $T = \{\tau_1, \dots, \tau_m\}$, where the transaction τ_i is related to the csQCA case $c_i = (x_i, y_i)$ in the following way: $\tau_i = \{\phi_{i1}, \phi_{i2}, \dots, \phi_{in}, \eta_i\}$, where

$$\phi_{ij} = \begin{cases} V_j & \text{if } x_{ij} = 1 \\ v_j & \text{if } x_{ij} = 0, \end{cases} \quad \eta_i = \begin{cases} R & \text{if } y_i = 1 \\ r & \text{if } y_i = 0, \end{cases}$$

Thus, each transaction τ_i can be regarded as a collection of n elements, which are either the variables V_j or their negation v_j , together with the presence of the outcome R or its absence r .

Remark 2. It is important to note that, although there are $2n + 2$ items in I , the conditions imposed on the set of transactions T greatly limit the range of possibilities: first, all transactions are of length $n + 1$; second, no transaction can contain a variable and its negation (V_j and v_j or R and r).

Now, we need to establish the relationship between the k -configurations in the csQCA model and the association rules in the AR model.

Let $S_k = (\mathcal{V}_k, \mu)$ be a k -configuration conducive to the outcome R . Let us assume that the set $\mathcal{V}_k = \{V_{j1}, \dots, V_{jk}\}$; then, S_k is equivalent to the association rule $X \Rightarrow R$, where

$$X = \{\psi_1, \dots, \psi_k\},$$

is

$$\psi_l = \begin{cases} V_{jl} & \text{if } \mu(V_{jl}) = 1 \\ v_{jl} & \text{if } \mu(V_{jl}) = 0. \end{cases}$$

Remark 3. We use $\mathcal{T}$ to denote mapping between the csQCA k -configurations and the association rules; that is,

$$\mathcal{T}(S_k) = (X \Rightarrow R).$$

Similarly, abusing the notation, we write $\mathcal{T}(\mathcal{C})$ to denote the element-wise transformation of the configurations defining $\mathcal{C}$ (see Definition 7).

With this relationship between k -configurations and association rules, it is obvious that we have established a one-to-one correspondence between the cases described in the dataset matrix M and the set of transactions T . Moreover, we verify that if the case c_i is compatible with the configuration S_k , then $X \subset \tau_i$ is straightforward. That is, the antecedent of the association rule associated with S_k is included in the transaction τ_i associated with the case c_i .

However, it is important to note that in a csQCA problem, it is sufficient that there is a case in M compatible with a given k -configuration for it to be incorporated into the

solution set. Still, in an association rule problem, only the interesting rules are considered in the solution set, and those depend on the scores of support and confidence. Therefore, we need to compare the metrics in QCA and ARM.

Given the consistency and coverage definitions (Equations (3) and (5)) and the above definition of the item and transaction sets, it is clear that

$$n_x = |\{\tau \in T \mid X \subseteq \tau\}|, \quad n_y = |\{\tau \in T \mid Y \subseteq \tau\}|, \quad n_{xy} = |\{\tau \in T \mid X \cup Y \subseteq \tau\}|.$$

Therefore, given a k -configuration S_k and its equivalent association rule $X \Rightarrow R$, we first see that the quantity n_{xy} is directly related to the support of the rule $X \Rightarrow R$. In particular,

$$\text{supp}(X \Rightarrow R) = \frac{n_{xy}}{m}, \quad (7)$$

where m denotes the number of rows (cases) in the sample data matrix M . In addition, there is a direct relationship between the consistency of the configuration and the confidence of the equivalent rule; that is,

$$\text{cons}(S_k) = \text{conf}(X \Rightarrow R) \quad (8)$$

On the other hand, the consistency can be regarded as the confidence of the necessary condition for the outcome; that is,

$$\text{cov}(S_k) = \text{conf}(R \Rightarrow X) \quad (9)$$

With these considerations in mind, the following theorem relates the solutions of a csQCA problem to the mining of a certain type of association rules.

Theorem 1. Let $(\mathcal{V}, R, M)$ be a csQCA model, and let (I, T) be the association rule mining problem defined above in this section. Then,

$$\mathcal{T}(\mathcal{C}) \subseteq \mathcal{AR},$$

where $\mathcal{AR}$ denotes the set of all interesting rules with minimum support and confidence thresholds given by

$$\text{minsupp} = \frac{1}{m} \quad \text{minconf} = 1.$$

Proof. Let $S_k \in \mathcal{C}$, and $\mathcal{T}(S_k) = X \Rightarrow R$ is its equivalent association rule. Since the configuration $S_k = (\mathcal{V}, \mu)$ is conducive to outcome R , there exists at least one case $\mathbf{c}_i = (\mathbf{x}_i, y_i) \in M$ such that $y_i = 1$, and $\mu(V_j) = x_{ij}$ for all $V_j \in \mathcal{V}$. Therefore, $n_{xy} \geq 1$. Then, from Equation (7),

$$\text{supp}(X \Rightarrow R) = \frac{n_{xy}}{m} \geq \frac{1}{m}.$$

On the other hand, if the csQCA model $(\mathcal{V}, R, M)$ is consistent, then $\text{cons}(S_k) = 1$, so from (8), $\text{conf}(X \Rightarrow R) = 1$.

Hence, association rule $X \Rightarrow R$ is an interesting association rule with the thresholds

$$\text{minsupp} = \frac{1}{m}, \quad \text{and} \quad \text{minconf} = 1,$$

so $(X \Rightarrow R) \in \mathcal{AR}$, and then $\mathcal{T}(\mathcal{C}) \subseteq \mathcal{AR}$. $\square$

Remark 4. It is important to note that of the two thresholds set in the theorem, the first is set to $1/m$ such that it is sufficient for one case to exist for it to be included in the solution set of the

csQCA problem. The second is 1 because it is assumed that the matrix M is consistent and, therefore, there are no contradictions.

These two conditions, which may seem very restrictive initially, have been commonly used in csQCA since its origin. If, for some reason, a larger number of repetitions were required in the csQCA problem for a case to be considered conducive to the result, it would suffice to change the minsupp threshold accordingly.

Likewise, although not common in csQCA practice, if a certain degree of inconsistency in the cases was allowed, it would be sufficient to lower the confidence threshold to the desired consistency value.

3.4. Reducible Association Rules

One of the main issues with association rules is the number of solutions they generate. Khedr et al., 2012 [50], state that many solutions generated by association rule algorithms may be redundant and irrelevant. Therefore, reducing the number of rules is one of the primary objectives of research on association rule extraction. With this in mind, the main goal of this section is to devise an association rule minimization algorithm for later comparison with the solutions produced by the QCA algorithms.

In the particular case of association rules stemming from a QCA problem, such as those defined in the previous section, we demonstrate a result below that will allow us to simplify association rules simply and efficiently, yielding the same results as the Quine–McCluskey reduction seen for csQCA.

Theorem 2. Let $X_1 \Rightarrow R$ and $X_2 \Rightarrow R$ be two interesting association rules such that their corresponding k -configurations, $S_k^{(1)}$ and $S_k^{(2)}$, are reducible to the $(k-1)$ -configuration $S_{k-1}^{(*)}$. Then, the association rule corresponding to the reduced configuration $S_{k-1}^{(*)}$ is also an interesting rule.

Proof. If $X_1 \Rightarrow R$ and $X_2 \Rightarrow R$ are interesting rules, then both rules meet the minimum support and minimum confidence criteria:

$$\text{supp}(X_i \Rightarrow R) \geq \text{min supp}, \quad \text{conf}(X_i \Rightarrow R) \geq \text{min conf}, \quad i = 1, 2. \quad (10)$$

Let $X^* \Rightarrow R$ denote the association rule corresponding to the configuration $S_{k-1}^{(*)}$ (i.e., $\mathcal{T}(S_{k-1}^{(*)}) = (X^* \Rightarrow R)$). To prove that $X^* \Rightarrow R$ is also an interesting rule, we need to show that this rule also verifies the minimum support and confidence criteria.

Since $S_k^{(1)} = (\mathcal{V}_k^{(1)}, \mu_1)$ and $S_k^{(2)} = (\mathcal{V}_k^{(2)}, \mu_2)$ are reducible to the $(k-1)$ -configuration $S_{k-1}^{(*)} = (\mathcal{V}^{(*)}, \mu_*)$, according to Definition 6, we can consider, without any loss of generalization, that $\mathcal{V}_k^{(1)} = \mathcal{V}_k^{(2)} = \{V_1, \dots, V_k\}$, and there exists a variable that we may assume is V_k , such that

$$\mathcal{V}^{(*)} = \{V_1, \dots, V_{k-1}\},$$

and

$$\mu_*(V_i) = \mu_1(V_i) = \mu_2(V_i), \quad i = 1, \dots, k-1.$$

Moreover, since the variable V_k is the only one in which the configurations $S_k^{(1)}$ and $S_k^{(2)}$ differ, we may assume that, for example,

$$\mu_1(V_k) = 1, \quad \mu_2(V_k) = 0.$$

After these considerations, the association rules related to these configurations, $X_1 \Rightarrow R$, $X_2 \Rightarrow R$, and $X^* \Rightarrow R$, are given by

$$X_1 = \{\psi_1, \dots, \psi_{k-1}, V_k\}, \quad X_2 = \{\psi_1, \dots, \psi_{k-1}, v_k\}, \quad X^* = \{\psi_1, \dots, \psi_{k-1}\},$$

where

$$\psi_i = \begin{cases} V_i & \text{if } \mu_1(V_i) = 1 \\ v_i & \text{if } \mu_1(V_i) = 0, \end{cases}$$

for $i = 1, \dots, k-1$. Thus, $X_1 = X^* \cup \{V_k\}$ and $X_2 = X^* \cup \{v_k\}$.

Now, we show that any transaction containing X^* either contains X_1 or X_2 ; that is,

$$\{\tau \in T : X^* \subset \tau\} = \{\tau \in T : X_1 \subset \tau\} \cup \{\tau \in T : X_2 \subset \tau\} = T_1 \cup T_2.$$

For this purpose, we use the double inclusion method.

$\subseteq$ Let $\tilde{\tau} \in \{\tau \in T : X^* \subset \tau\}$; then, from Remark 2, we know that all transactions have the length $n+1$ and that they contain each variable once, either in its affirmative (V_i) or negative (v_i) form. Therefore, since $X^* \subset \tilde{\tau}$, it is ensured that either

$$\tilde{\tau} = \{\psi_1, \dots, \psi_{k-1}, V_k, \dots\}, \quad \text{or} \quad \tilde{\tau} = \{\psi_1, \dots, \psi_{k-1}, v_k, \dots\}.$$

In either case, $\tilde{\tau} \in T_1 \cup T_2$.

$\supseteq$ Let $\tilde{\tau} \in T_1 \cup T_2$. If $\tilde{\tau} \in T_1$, then $\tilde{\tau} = \{\psi_1, \dots, \psi_{k-1}, V_k, \dots\}$, and if $\tilde{\tau} \in T_2$, then $\tilde{\tau} = \{\psi_1, \dots, \psi_{k-1}, v_k, \dots\}$. In either case, $X^* = \{\psi_1, \dots, \psi_{k-1}\} \subset \tilde{\tau}$, and therefore, $\tilde{\tau} \in \{\tau \in T : X^* \subset \tau\}$.

Moreover, it is trivial that T_1 and T_2 are disjoint sets as, if a transaction contains X_1 , then it contains the affirmative variable V_k and, thus, cannot contain X_2 . Therefore,

$$n_{x^*} = |\{\tau \in T : X^* \subset \tau\}| = |T_1 \cup T_2| = |T_1| + |T_2| = n_{x_1} + n_{x_2}.$$

Analogously, it can also be shown that

$$n_{x^*y} = n_{x_1y} + n_{x_2y}.$$

Therefore,

$$\text{supp}(X^* \Rightarrow R) = \frac{n_{x^*}}{m} = \frac{n_{x_1} + n_{x_2}}{m} = \text{supp}(X_1 \Rightarrow R) + \text{supp}(X_2 \Rightarrow R), \quad (11)$$

and

$$\text{conf}(X^* \Rightarrow R) = \frac{n_{x^*y}}{n_{x^*}} = \frac{n_{x_1y} + n_{x_2y}}{n_{x_1} + n_{x_2}}. \quad (12)$$

We are now in a position to prove that $X^* \Rightarrow R$ is an interesting rule. From Equations (10) and (11), it is obvious that

$$\text{supp}(X^* \Rightarrow R) \geq 2 \min \text{supp} > \min \text{supp}.$$

On the other hand, considering the algebraic property that states that if $a, b, c, d > 0$ and $a/b \leq c/d$, then

$$\frac{a}{b} \leq \frac{a+c}{b+d} \leq \frac{c}{d},$$

and if we assume, without any loss of generalization, that $n_{x_1y}/n_{x_1} \leq n_{x_2y}/n_{x_2}$, then from (12),

$$\frac{n_{x_1y}}{n_{x_1}} \leq \frac{n_{x_1y} + n_{x_2y}}{n_{x_1} + n_{x_2}} \leq \frac{n_{x_2y}}{n_{x_2}}.$$

Thus,

$$\text{conf}(X_1 \Rightarrow R) \leq \text{conf}(X^* \Rightarrow R) \leq \text{conf}(X_2 \Rightarrow R),$$

and, from (10), we can conclude that

$$\text{conf}(X^* \Rightarrow R) \geq \text{conf}(X_1 \Rightarrow R) \geq \min \text{conf},$$

and, therefore, $X^* \Rightarrow R$ is an interesting rule. $\square$

This result allows us to obtain a method to minimize association rules in the Quine–McCluskey sense, but with the advantage of not needing to use truth tables (whose size grows exponentially with the number of variables). The Apriori method provides a list of interesting association rules that we know contain all the solutions to the csQCA problem. Now, as we know from Theorem 2, if two interesting rules are reducible, their reduction is also interesting and, therefore, it will already be in the list. Hence, we only have to eliminate reducible rules from the list.

4. Examples

4.1. Internet Blockades in Election Times

In this initial example, the csQCA and ARM methodologies are illustrated using a case study derived from [51], who analyzed the presence or absence of Internet shutdowns during electoral periods, including the days preceding the elections, the election day itself, and the subsequent days. The data used, as presented in Table 2, were obtained from a comparative study encompassing 33 countries within the Sub-Saharan African region (treated as cases/transactions). Consequently, for educational purposes, this study explores three causal conditions (ISP, AT, and EV) alongside an outcome (IS).

Table 2. A summary of cases and variables.

N.	Case (Country, Election Year)	ISP State Ownership	Autocracy	Electoral Violence	Internet Shutdown
1	Benin, 2015	1	0	0	0
2	Benin, 2016	1	0	0	0
3	Botswana, 2014	1	0	0	0
4	Burkina Faso, 2015	0	0	0	0
5	Burundi, 2015	1	1	1	1
6	CAR, 2015	0	1	0	0
7	CAR, 2016	0	0	0	0
8	Chad, 2016	1	1	1	1
9	Djibouti, 2016	0	1	1	1
10	Equatorial Guinea, 2016	1	0	0	0
11	Ethiopia, 2015	1	1	0	1
12	Gabon, 2016	1	1	0	1
13	Gambia, 2016	0	0	1	1
14	Ghana, 2016	1	1	0	1
15	Guinea, 2015	0	0	0	0
16	Guinea-Bissau, 2014	1	0	0	0
17	Ivory Coast, 2015	0	0	0	0
18	Ivory Coast, 2016	0	0	0	0
19	Lesotho, 2015	1	0	0	0
20	Malawi, 2014	0	0	0	0
21	Mauritania, 2014	0	0	0	0
22	Mozambique, 2014	0	1	0	0

Table 2. Cont.

N.	Case (Country, Election Year)	ISP State Ownership	Autocracy	Electoral Violence	Internet Shutdown
23	Namibia, 2014	1	0	1	0
24	Niger, 2016	1	0	0	0
25	Nigeria, 2015	1	0	0	0
26	Republic of Congo, 2016	1	0	0	0
27	South Africa	1	0	0	0
28	Sudan, North, 2015	0	1	1	1
29	Tanzania, 2015	1	0	1	0
30	Togo, 2015	1	1	0	1
31	Uganda, 2016	0	1	1	1
32	Zambia, 2015	1	0	0	0
33	Zambia, 2016	1	0	1	0

4.1.1. csQCA Approach

Following the notation introduced in Section 3.1, the csQCA model is given by $(\mathcal{V}, R, M)$, where

- $\mathcal{V} = \{\text{ISP}, \text{AT}, \text{EV}\}$ is the variable set. More precisely, ISP refers to “State-Owned Internet Service Providers” (ISPs). These ISPs play a crucial role, especially in contexts where the government has direct control over the telecommunications infrastructure, such as the national Internet backbone, allowing it to exert significant influence over Internet access and availability. AT is the variable “Autocracy”. It is used to categorize the political regime of a country, distinguishing between “1” if the regime is an autocracy and “0” if it is a democracy. EV refers to the occurrence of violence associated with the electoral process.
- $R = \text{IS}$. The outcome variable is “Internet Shutdown”.
- M is the data matrix given in Table 2.

Table 3 presents the truth table for the occurrence of Internet shutdowns. In the present study, with three causal conditions (variables) considered, there are $2^3 = 8$ theoretically possible configurations across the 33 cases examined. While all eight of these configurations appear in the empirical data, achieving such empirical completeness is not typical. When certain configurations lack empirical representation, it results in limited diversity, a frequently encountered constraint in small- and medium-scale comparative studies.

Next, we perform the so-called sufficiency analysis, whose central objective is to evaluate all possible combinations of causal conditions and identify which constitute consistent subsets of the outcome [22]. In this situation, the configuration is considered sufficient for the outcome (IS) [38]. Analyzing Table 3, we can observe four 3-configurations conducive to the IS outcome; namely,

$$\begin{aligned}
 S_3^{(1)} &= \begin{bmatrix} \text{ISP} & \text{AT} & \text{EV} \\ 0 & 0 & 1 \end{bmatrix} & S_3^{(2)} &= \begin{bmatrix} \text{ISP} & \text{AT} & \text{EV} \\ 0 & 1 & 1 \end{bmatrix} \\
 S_3^{(3)} &= \begin{bmatrix} \text{ISP} & \text{AT} & \text{EV} \\ 1 & 1 & 0 \end{bmatrix} & S_3^{(4)} &= \begin{bmatrix} \text{ISP} & \text{AT} & \text{EV} \\ 1 & 1 & 1 \end{bmatrix}
 \end{aligned}$$

Table 3. Truth table for the outcome “Internet Shutdown”.

ISP	AT	EV	IS	N.Cases	Const. ¹	Cases
0	0	1	1	1	1	Gambia_16
0	1	1	1	3	1	Djibouti_16, Sudan, North_15, Uganda_16
1	1	0	1	4	1	Gabon_16, Ghana_16, Ethiopia_15, Togo_15
1	1	1	1	2	1	Burundi_15, Chad_16
0	0	0	0	7	0	Burkina Faso_15, CAR_15, Guinea_15, Ivory Coast_16, Malawi_14, Mauritania_14
0	1	0	0	2	0	CAR_15, Mozambique_14
1	0	0	0	11	0	Republic of Congo_16, Benin_15, Benin_16, Botswana_14, Lesotho_15, Guinea Bissau_16, Ivory Coast_16, Niger_16, Nigeria_15, Equatorial Guinea_16, Zambia_15, South Africa_14
1	0	1	0	3	0	Namibia_14, Tanzania_15, Zambia_16

¹ Consistency.

In this small-size example, it is easy to see that all 3-configurations are reducible to three 2-configurations. In particular,

$$\left. \begin{matrix} S_3^{(1)} \\ S_3^{(2)} \end{matrix} \right\} \rightarrow S_2^{(1)} = \begin{bmatrix} \text{ISP} & \text{EV} \\ 0 & 1 \end{bmatrix}, \quad \left. \begin{matrix} S_3^{(3)} \\ S_3^{(4)} \end{matrix} \right\} \rightarrow S_2^{(2)} = \begin{bmatrix} \text{ISP} & \text{AT} \\ 1 & 1 \end{bmatrix}.$$

$$\left. \begin{matrix} S_3^{(2)} \\ S_3^{(4)} \end{matrix} \right\} \rightarrow S_2^{(3)} = \begin{bmatrix} \text{AT} & \text{EV} \\ 1 & 1 \end{bmatrix}$$

Since the 2-configurations obtained are no longer reducible, the solution set of the QCA problem is

$$\mathcal{C} = \mathcal{C}_2 = \{S_2^{(1)}, S_2^{(2)}, S_2^{(3)}\}.$$

To interpret the solution obtained, we follow the standard notation in QCA, in which variables are denoted in upper and lower case to indicate the presence and absence of a causal condition, respectively. The symbol “*” is used to aggregate variables, and the “+” sign denotes the union of configurations.

Hence, with this notation, the 2-configurations obtained can be written as

$$S_2^{(1)} = isp * EV, \quad S_2^{(2)} = ISP * AT, \quad S_2^{(3)} = AT * EV,$$

and therefore, the minimization of the truth table for the Internet shutdown leads to the following solution:

$$\mathcal{C} : isp * EV + ISP * AT + AT * EV \rightarrow IS.$$

In summary, three solutions for Internet shutdowns during recent elections in Sub-Saharan African (SSA) countries were identified. In the first scenario, the ISP ownership is not public, and significant incidents of electoral violence occurred. In the second scenario, an autocratic state that holds majority ownership of Internet service providers (ISPs) initiates a disruption of the Internet service during elections. Finally, the combination of an authoritarian regime with episodes of electoral violence also led to disruptions in Internet access, regardless of public or private ISP ownership.

Regarding the metrics, it is worth noting that, as shown in Table 4, all three solution configurations have a consistency of 1, as there are no contradictory cases. The coverage of the cases is 0.4, 0.6, and 0.5, showing that when an Internet service disruption occurs,

the second causal configuration ($ISP * AT$) is the most frequent (examples of this can be observed in events in 2015 (Ethiopia, Togo, and Burundi) and 2016 (Gabon, Ghana, and Chad)).

Table 4. A table of solutions with consistency and coverage.

Solution	X_i	Y_i	n_x	n_y	n_{xy}	Consistency	Coverage
$isp * EV \rightarrow IS$	$isp * EV$	IS	4	10	4	1	0.4
$ISP * AT \rightarrow IS$	$ISP * AT$	IS	6	10	6	1	0.6
$AT * EV \rightarrow IS$	$AT * EV$	IS	5	10	5	1	0.5

4.1.2. Positive and Negative Association Rule Solution

From the viewpoint of association rule mining, for this problem, we define the set of items as

$$I = \{ISP, isp, AT, at, EV, ev, IS, is\}, \quad (13)$$

and the set of transactions T is defined by the rows in Table 2. For example, row 1, corresponding to the 2015 electoral process in Benin, indicates the transaction $\tau = \{ISP, at, ev, is\}$.

The Apriori algorithm for mining positive and negative boolean association rules follows four main steps [52].

- Generate Candidate Item Sets as Subsets of the Universal Item Set I given in (13): Start with individual items (1 item set), then iteratively combine frequent item sets from the previous step to form larger candidate item sets (k item sets).
- Prune Using Minimum Support: Eliminate candidates that do not meet a user-defined minimum support threshold (frequency in the dataset). This relies on the **Apriori principle**: any subset of a frequent item set must itself be frequent.
- Repeat: Continue to generate and prune item sets of increasing size until no new frequent item sets are found.
- Form Association Rules: From the final frequent item sets, generate rules (e.g., $X \rightarrow Y$) and filter them using a minimum confidence threshold (how often Y appears when X is present).

If C_k denotes the candidate item set of length k , then it can be shown that, if there are n variables and an outcome, then the number of items in C_k (taking into account that C_k cannot contain a variable and its negation) is given by

$$|C_k| = \frac{\prod_{i=0}^{k-1} 2(n+1-i)}{k!}, \quad k = 1, 2, \dots, n+1.$$

Therefore, in this example, $|C_1| = 8$, $|C_2| = 24$, $|C_3| = 32$, and $|C_4| = 16$.

As previously stated, Apriori is a bottom-up algorithm that starts with the smallest sets and increases in size, eliminating combinations that do not exceed a minimum support threshold. In our case, and following Theorem 1, the minimum support threshold is

$$\min \text{supp} = \frac{1}{m} = \frac{1}{33} \approx 0.03.$$

It is important to note that association rule mining does not initially consider some items as antecedents and others as consequents in the same way that QCA separates the antecedent variables and the outcome to be analyzed. Consequently, the preliminary step involves extracting the item sets containing the response variable ' IS ', as well as analyzing which of these item sets meet the minimum support and confidence conditions.

Table 5 shows the item sets that have passed the minimum support threshold, the association rule they define, and the confidence of the rule. Association rules that have confidence 1 are highlighted in boldface.

Table 5. The association rules leading to the outcome “IS” and meeting the minimum support condition.

Item Set	Count	Association Rule	Count Ant.	Supp.	Conf.
{“isp”, “IS”}	4	isp $\rightarrow$ IS	13	0.12	0.31
{“at”, “IS”}	1	at $\rightarrow$ IS	22	0.03	0.05
{“ev”, “IS”}	4	ev $\rightarrow$ IS	24	0.12	0.17
{“ISP”, “IS”}	6	ISP $\rightarrow$ IS	20	0.18	0.30
{“AT”, “IS”}	9	AT $\rightarrow$ IS	11	0.27	0.82
{“EV”, “IS”}	6	EV $\rightarrow$ IS	9	0.18	0.67
{“isp”, “EV”, “IS”}	4	isp * EV $\rightarrow$ IS	4	0.12	1.00
{“ISP”, “EV”, “IS”}	2	ISP * EV $\rightarrow$ IS	5	0.06	0.40
{“ISP”, “ev”, “IS”}	4	ISP * EV $\rightarrow$ IS	15	0.12	0.27
{“ISP”, “AT”, “IS”}	6	ISP * AT $\rightarrow$ IS	6	0.18	1.00
{“isp”, “AT”, “IS”}	3	isp * AT $\rightarrow$ IS	5	0.09	0.60
{“AT”, “ev”, “IS”}	4	AT * ev $\rightarrow$ IS	6	0.12	0.67
{“AT”, “EV”, “IS”}	5	AT * EV $\rightarrow$ IS	5	0.15	1.00
{“at”, “EV”, “IS”}	1	AT * ev $\rightarrow$ IS	4	0.03	0.25
{“isp”, “at”, “IS”}	1	AT * ev $\rightarrow$ IS	8	0.03	0.13
{“isp”, “at”, “EV”, “IS”}	1	isp * at * EV $\rightarrow$ IS	1	0.03	1.00
{“isp”, “AT”, “EV”, “IS”}	3	isp * AT * EV $\rightarrow$ IS	3	0.09	1.00
{“ISP”, “AT”, “ev”, “IS”}	4	ISP * AT * ev $\rightarrow$ IS	4	0.12	1.00
{“ISP”, “AT”, “EV”, “IS”}	2	ISP * AT * EV $\rightarrow$ IS	2	0.06	1.00

Given the results shown in Table 5, the set $\mathcal{AR}$ of interesting association rules leading to the outcome “Internet Shutdown” is formed by the following rules:

$$r_1 : \text{isp} * \text{EV} \rightarrow \text{IS}$$

$$r_2 : \text{ISP} * \text{AT} \rightarrow \text{IS}$$

$$r_3 : \text{AT} * \text{EV} \rightarrow \text{IS}$$

$$r_4 : \text{isp} * \text{at} * \text{EV} \rightarrow \text{IS}$$

$$r_5 : \text{isp} * \text{AT} * \text{EV} \rightarrow \text{IS}$$

$$r_6 : \text{ISP} * \text{AT} * \text{ev} \rightarrow \text{IS}$$

$$r_7 : \text{ISP} * \text{AT} * \text{EV} \rightarrow \text{IS}.$$

It is evident that a relationship can be established between the causal configurations of the QCA problem and the association rules obtained. Indeed, recalling the definition of the operator $\mathcal{T}$ in Remark 3,

$$\mathcal{T}(\mathcal{C}_2) = \mathcal{T}\left(\left\{\mathcal{S}_2^{(1)}, \mathcal{S}_2^{(2)}, \mathcal{S}_2^{(3)}\right\}\right) = \{r_1, r_2, r_3\}$$

$$\mathcal{T}(\mathcal{C}_3) = \mathcal{T}\left(\left\{\mathcal{S}_3^{(1)}, \mathcal{S}_3^{(2)}, \mathcal{S}_3^{(3)}, \mathcal{S}_4^{(3)}\right\}\right) = \{r_4, r_5, r_6, r_7\}$$

Consequently, the solution set of the QCA problem is contained in AR. Furthermore, it should be noted that the set AR encompasses both the initial and the minimized causal configurations. To obtain the same result, it suffices to determine r_1 , r_2 , and r_3 from among the rules and whether they can be obtained by minimizing the rules $r_4, \dots, r_7$. It is clear that this is the case, and that if we were to complete this minimization, we would obtain the same set of solutions.

4.2. Opposition to Immigration in Europe

In the following case study, we show how the association rule methodology can also be used efficiently when the number of cases is larger. To achieve this, we adapt the data from [53] to present another comparison between QCA and association rules. These data consist of a robust sample of 2223 calibrated observations from a study investigating opposition to immigration in Europe. The original data were collected using a 10-point ordinal scale, which allowed participants to express their levels of agreement and disagreement on the issue of immigration. This dataset covers a total of six variables, five of which represent potential conditioning factors: 1—low education [LOWEDU], 2—preference for cultural unity [UNITY], 3—a lack of immigrant friends [NOFRIEND], 4—perceived economic threat [THREAT], and 5—gender [MAN]. In addition to these, there is an outcome variable called [OPPOSITION]. Table 6 summarizes the sample used in this study, highlighting the frequencies corresponding to each possible combination of participants' responses. Because we have five independent variables and two answer options (yes or no) for each, there are 32 possible combinations of answers, as shown in Table 6. For example, in the first line of the table, we see that 22 individuals responded negatively to the first variable and affirmatively to the remaining four variables.

Table 6. Summary data for case study 2: opposition to immigration.

	LOWEDU	UNITY	NOFRIEND	THREAT	MAN	OPPOSITION	n
1	0	1	1	1	1	1	22
2	1	0	1	0	1	1	47
3	1	0	1	1	0	1	163
4	1	0	1	1	1	1	64
5	1	1	0	0	0	1	41
6	1	1	0	0	1	1	24
7	1	1	0	1	0	1	98
8	1	1	0	1	1	1	76
9	1	1	1	0	0	1	83
10	1	1	1	0	1	1	40
11	1	1	1	1	0	1	196
12	1	1	1	1	1	1	193
13	0	0	0	0	0	0	93
14	0	0	0	0	1	0	123
15	0	0	0	1	0	0	26
16	0	0	0	1	1	0	40
17	0	0	1	0	0	0	56
18	0	0	1	0	1	0	47
19	0	0	1	1	0	0	19
20	0	0	1	1	1	0	30
21	0	1	0	0	0	0	22
22	0	1	0	0	1	0	37
23	0	1	0	1	0	0	5
24	0	1	0	1	1	0	17
25	0	1	1	0	0	0	21
26	0	1	1	0	1	0	49
27	0	1	1	1	0	0	7
28	1	0	0	0	0	0	101
29	1	0	0	0	1	0	59
30	1	0	0	1	0	0	172
31	1	0	0	1	1	0	156
32	1	0	1	0	0	0	96

4.2.1. csQCA Solution

In this context, we present the solutions obtained with the csQCA algorithm to analyze opposition to immigration in Europe. Using the statistical programming language R

(v. 4.3.0) [54] and the QCA package (v. 3.23) [8], we found that these solutions comprise four causal combinations that lead to the analyzed phenomenon:

1. LOWEDU * UNITY.
2. LOWEDU * NOFRIEND * THREAT.
3. LOWEDU * NOFRIEND * MAN.
4. UNITY * NOFRIEND * THREAT * MAN.

It can be seen that the most concise solution, characterized by the smallest number of causal conditions, is the first, comprising just two conditions: low schooling and preference for cultural unity. On the other hand, the solution with the greatest number of causal conditions is the fourth, which comprises four different causal variables.

4.2.2. Positive and Negative Association Rule Solution

Next, we transform the csQCA problem into a binary association rule mining problem, as was accomplished in the previous section. To achieve this goal, the item set $\mathcal{I}$ was constructed, which contains 12 elements (5 variables and 1 outcome and its negations). The list of transactions T was obtained from the truth table given in Table 6, and the arules package [55,56] was used to determine the interesting association rules, i.e., those with a support higher than 0.0004 ($\approx 1/2263$) and a confidence of 1.

Table 7 shows the solution set of association rules derived from the proposed problem, together with the corresponding metrics. A total of 39 rules were generated, organized in increasing order of complexity. The least complex rule combines just two variables, while the most complex rules combine five different variables.

Table 7. Association rules with 100% confidence (Conf = 1) implying {OPPOSITION}.

	LHS → OPPOSITION	Support	Coverage	Count
1	LOWEDU * UNITY	0.34	0.34	751
2	LOWEDU * threat * UNITY	0.08	0.08	188
3	LOWEDU * MAN * UNITY	0.15	0.15	333
4	LOWEDU * nofriend * UNITY	0.11	0.11	239
5	LOWEDU * NOFRIEND * UNITY	0.23	0.23	512
6	LOWEDU * man * UNITY	0.19	0.19	418
7	LOWEDU * THREAT * UNITY	0.25	0.25	563
8	LOWEDU * MAN * NOFRIEND	0.15	0.15	344
9	LOWEDU * NOFRIEND * THREAT	0.28	0.28	616
10	LOWEDU * MAN * threat * UNITY	0.03	0.03	64
11	LOWEDU * nofriend * threat * UNITY	0.03	0.03	65
12	LOWEDU * NOFRIEND * threat * UNITY	0.06	0.06	123
13	LOWEDU * man * threat * UNITY	0.06	0.06	124
14	LOWEDU * MAN * nofriend * UNITY	0.04	0.04	100
15	MAN * NOFRIEND * THREAT * UNITY	0.10	0.10	215
16	LOWEDU * MAN * NOFRIEND * UNITY	0.10	0.10	233
17	LOWEDU * MAN * THREAT * UNITY	0.12	0.12	269
18	LOWEDU * man * nofriend * UNITY	0.06	0.06	139
19	LOWEDU * nofriend * THREAT * UNITY	0.08	0.08	174
20	LOWEDU * man * NOFRIEND * UNITY	0.13	0.13	279
21	LOWEDU * NOFRIEND * THREAT * UNITY	0.17	0.17	389
22	LOWEDU * man * THREAT * UNITY	0.13	0.13	294
23	LOWEDU * MAN * NOFRIEND * threat	0.04	0.04	87
24	LOWEDU * MAN * NOFRIEND * THREAT	0.12	0.12	257
25	LOWEDU * MAN * NOFRIEND * unity	0.05	0.05	111
26	LOWEDU * man * NOFRIEND * THREAT	0.16	0.16	359
27	LOWEDU * NOFRIEND * THREAT * unity	0.10	0.10	227

Table 7. Cont.

	LHS → OPPOSITION	Support	Coverage	Count
28	lowedu * MAN * NOFRIEND * THREAT * UNITY	0.01	0.01	22
29	LOWEDU * MAN * nofriend * threat * UNITY	0.01	0.01	24
30	LOWEDU * MAN * NOFRIEND * threat * UNITY	0.02	0.02	40
31	LOWEDU * man * nofriend * threat * UNITY	0.02	0.02	41
32	LOWEDU * man * NOFRIEND * threat * UNITY	0.04	0.04	83
33	LOWEDU * MAN * nofriend * THREAT * UNITY	0.03	0.03	76
34	LOWEDU * MAN * NOFRIEND * THREAT * UNITY	0.09	0.09	193
35	LOWEDU * man * nofriend * THREAT * UNITY	0.04	0.04	98
36	LOWEDU * man * NOFRIEND * THREAT * UNITY	0.09	0.09	196
37	LOWEDU * MAN * NOFRIEND * threat * unity	0.02	0.02	47
38	LOWEDU * MAN * NOFRIEND * THREAT * unity	0.03	0.03	64
39	LOWEDU * man * NOFRIEND * THREAT * unity	0.07	0.07	163

However, thanks to Theorem 2, we know that if two interesting association rules are reducible, then the reduced rule is also interesting. This insight enables a very simple minimization procedure: we can eliminate from the list all rules for which there exists another that differs from it in only one bit. For example, rules 2 and 7, 3 and 6, and 4 and 5 are all reducible to rule 1. Similarly, all rules of length 5 (i.e., rules 28 to 39) are reducible to rules with a length of 4, which are already present in the list, and so on.

After filtering out all reducible rules, the remaining rules are as follows:

1. Rule 1: LOWEDU * UNITY.
2. Rule 8: LOWEDU * MAN * NOFRIEND.
3. Rule 9: LOWEDU * NOFRIEND * THREAT.
4. Rule 15: MAN * NOFRIEND * THREAT * UNITY.

These coincide with the solution obtained for the csQCA problem in the previous subsection.

4.3. Numerical Experiments

The computational cost of a truth table analysis and Quine-McCluskey minimization has been widely discussed in the QCA literature. It is well established that QCA studies typically involve a limited number of variables (rarely exceeding ten) and a relatively small number of cases (often only in the tens). Examples involving hundreds or thousands of cases—such as that shown in the earlier example—are exceptionally rare.

Initially, QCA analysis methods were especially applicable to studies with an intermediate number of cases (between 10 and 50 [9]). Over the years, several researchers have developed studies handling thousands of cases for different purposes [23,37,57]. In fact, in theory, the sample size is only limited by hardware and software limitations, and not necessarily by the limitation imposed by the methodology itself.

We conducted a series of numerical experiments to illustrate the practical differences between csQCA and association rule mining, focusing particularly on computational feasibility. These experiments simulate problems with varying numbers of variables and cases, allowing for a systematic comparison of the scalability and efficiency of both approaches. For each scenario, we randomly generated consistent data tables (i.e., without contradictions) with a fixed number of variables and cases. Each experiment was repeated five times, and we recorded the mean computation time and memory usage for the csQCA algorithm (implemented via the QCA R package [8]), the Apriori rule mining algorithm (using the arules R package [55,56]), and a custom implementation of an association rule minimization algorithm in R. All computations were conducted on a 13th Gen Intel(R) Core(TM) i7-1360P (2.20 GHz) with 16 GB of RAM and no use of GPU acceleration. The code for this exper-

iment is available at the GitHub repository <https://github.com/rbensua/csQCAARM/> (accessed on 30 April 2025).

In the first experiment, we set a small number of variables (five variables) and varied the number of cases from 5 to 10,000. The results shown in Figure 1 show that csQCA required almost a thousand times greater memory usage than Apriori. The minimization of association rules, on the other hand, required only about 300 Kb, regardless of the number of cases.

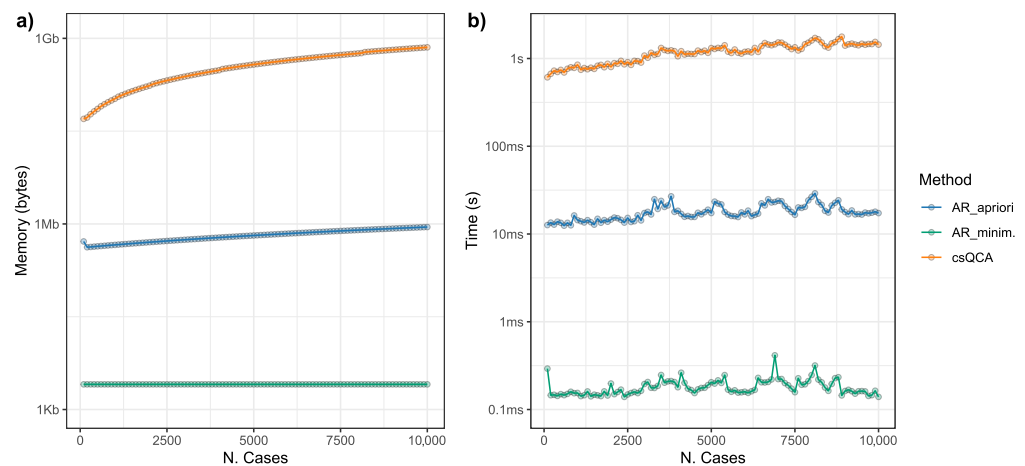


Figure 1. A comparison between csQCA and AR mining memory usage (a) and computation time (b) as the number of cases in the dataset increases. In both plots the scale in the y-axis is logarithmic. The number of variables was fixed at 5.

In the second experiment, we set the number of cases at 60 and varied the number of variables from 5 to 21. The results, shown in Figure 2, reveal that with both csQCA and AR, the memory usage and processing times both increase exponentially with the number of variables (note the logarithmic scales). However, within the range of a typical csQCA problem (5–15 variables), ARM memory usage—when both finding the rules via the Apriori algorithm and the minimization procedure—is significantly lower than that of csQCA. For a larger number of variables, the memory usage of Apriori rapidly increases (which is a well-known issue in association rule mining). Regarding computation times, AR algorithms performed significantly better than csQCA, which rapidly faces severe computational limitations as the number of variables increases (struggling to find solutions when dealing with more than 20 variables, as also noted by [30]). In the second experiment, we fixed the number of cases at 60 and varied the number of variables from 5 to 21. The results, shown in Figure 2, indicate that both the csQCA and AR methods exhibit exponential increases in memory usage and processing time as the number of variables grows (note the logarithmic scales on both axes).

Within the typical range of csQCA applications (5–15 variables), both AR steps (rule generation via the Apriori algorithm and the subsequent minimization procedure) use significantly less memory than csQCA. However, for larger numbers of variables, the memory usage of Apriori increases rapidly, which is a well-known limitation in association rule mining, giving values of the same order of magnitude as csQCA.

Regarding computation time, both AR algorithms also outperform csQCA. The csQCA computation times are in the range of [0.9 s–236 s], while the AR times combined are in the range of [0.02 s–105.7 s]. In this regard, csQCA shows a steep increase in processing time and becomes computationally infeasible for more than 20 variables, which is consistent with the findings of [30].

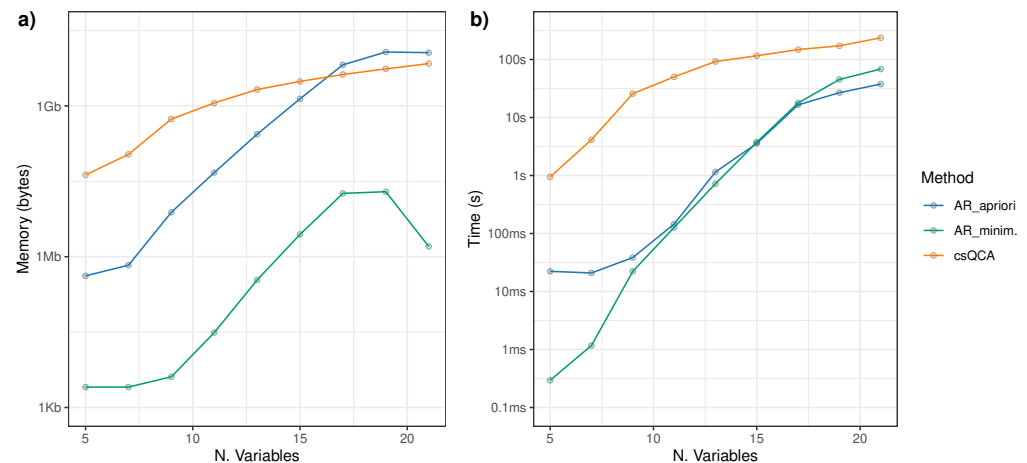


Figure 2. Comparison between csQCA and AR mining memory usage (a) and computation time (b) as the number of variables (conditions) increases. In both plots the scale in the y-axis is logarithmic. The number of cases was fixed at 60.

Remark 5. It is important to mention that although the solutions obtained with csQCA and ARM were exactly the same in the examples seen in the previous sections, in the numerical experiments, this was not the case. Obviously, according to Theorem 1, the solutions of csQCA were included in the list of interesting and irreducible rules of ARM, but ARM found more rules. This discrepancy arises from differences in the minimization algorithms used. In csQCA, the Quine–McCluskey algorithm follows a top-down approach, meaning that smaller configurations are always derived as simplifications of larger ones. In contrast, ARM starts by generating all potentially interesting rules and then eliminates those that are reducible. As a result, csQCA cannot produce configurations that do not originate from the minimization of more complex rules. We did not include the full list of solutions in the numerical experiments because, in higher-dimensionality cases, the number of resulting rules reached tens of millions.

5. Discussion

At first glance, both Qualitative Comparative Analysis (QCA) and association rule mining rest on case-based, Boolean “if-then” logic and seek to identify patterns of conditions or attributes that co-occur or link to a particular outcome [1,4]. However, their goals diverge: QCA explicitly engages with causal complexity, enabling multiple intersecting pathways to an outcome, whereas association rules capture statistical co-occurrence without implying causation [1,58].

The Quine–McCluskey minimization algorithm, which was originally devised to optimize logic circuits in electrical engineering [14], is applied by QCA in all its variants. This algorithm is employed to reduce complex configurations of conditions into simpler expressions. In contrast, the process of association rule mining starts with the enumeration of frequent item sets, subsequently leading to the derivation of rules based on support and confidence thresholds (see [4]).

Despite these conceptual and procedural differences, both techniques face the same combinatorial explosion: the search space for all possible condition combinations grows exponentially with the number of distinct attributes or conditions [4,7]. Furthermore, there is a formal parallel in the way rules are interpreted. In association rule mining, an implication $X \rightarrow Y$ is deemed strong if X satisfies a minimum frequency threshold, and every transaction containing X also includes Y [4]. Similarly, in QCA, the set of conditions X is considered necessary for outcome Y precisely when $Y \subseteq X$ across the observed cases [7], which may be regarded as an alternative formulation of the same concept.

This study provides a detailed comparative analysis of the solutions obtained using the association rule and Configurational Comparative Analysis (csQCA) methodologies.

Considering the evidence presented, the following key points emerge:

1. A one-to-one relationship can be established between a csQCA problem and a particular type of positive and negative binary association rule mining problem (Theorem 1). To our knowledge, this relationship has not been explored previously.
2. All the solutions generated by the csQCA algorithm were previously captured by the association rule mining algorithm. In other words, the emerging solution set from association rules acts as a superset of the emerging solutions from the csQCA approach. In this way, association rules show greater potential to generate a broader spectrum of solutions compared to the csQCA methodology.
3. The procedure used for obtaining solutions is completely different. While csQCA follows a top-down approach, in which the most complex solutions are obtained first, and, from the Quine–McCluskey minimization, simpler irreducible solutions are found, in ARM, a bottom-up approach is used (Apriori algorithm), starting with the simplest solutions and obtaining increasingly complex solutions. However, the Apriori algorithm does not simplify reducible solutions, since the objective for which it is designed is to obtain all frequent configurations. Despite this, we showed that it is not necessary to perform a Quine–McCluskey minimization since, as shown in Theorem 2, if two interesting rules are reducible, their reduction is also an interesting rule; thus, it is enough to filter and eliminate the reducible rules.
4. QCA originated in the context of social science research, where, traditionally, datasets have been relatively small in order to facilitate the extraction of interpretable solutions. In contrast, association rule mining emerged from the need to identify patterns within large-scale databases. As a result, the algorithms developed for ARM are specifically designed to operate efficiently on large datasets. Given the relationship established between QCA and association rule mining, this connection yields the possibility of using ARM techniques to obtain QCA-like solutions even when large datasets are involved, overcoming the traditional computational limitations of QCA.
5. Furthermore, it is important to emphasize that the solutions obtained through both methodologies exhibit identical robustness metrics, with confidence in association rule mining corresponding directly to consistency in csQCA. Additionally, all identified solutions adhere to an “if–then” structure, reinforcing their logical interpretability. Finally, both approaches are fundamentally grounded in set theory.

In this study, we show that association rules (ARs) produce solutions equivalent to those of csQCA. In contexts of limited data diversity, csQCA typically yields fewer solutions than ARs. However, the methods differ significantly. csQCA targets a single predefined outcome [59], while ARs can identify multiple outcomes and their complements [52], with outcomes either predefined or data-driven.

Both methods structure solutions as “if–then” statements joined by disjunctive (OR) operators [8]; however, only csQCA suggests causality, whereas AR identifies associations without causal claims [18]. Conditional elements must be selected in advance: causal conditions in csQCA [22] and items in AR [4]. Both approaches rely on Boolean structures and allow negative terms. Conjunctive (AND) combinations express conjunctural causality in csQCA [59], while in ARs, item sets are connected similarly but without causal interpretation, which must be inferred by the researcher.

Combinatorial generation also differs: csQCA yields 2^k combinations [9], whereas AR initially generates $3^k - 2^k + 1$ possibilities, usually filtered down to $2^k - 1$ [18].

6. Conclusions

In this study, we explored the theoretical and practical connections between Crisp-Set Qualitative Comparative Analysis (csQCA) and association rule mining (ARM). Using a formal mathematical framework, we demonstrated that every csQCA problem can be represented as a specific ARM problem involving positive and negative rules. By uncovering the internal mechanics of csQCA through this mathematical lens, we pave the way for a deeper understanding of its computational properties and compatibility with ARM approaches.

We showed that key QCA concepts such as consistency and coverage naturally correspond to ARM metrics such as confidence and lift. This equivalence was further supported by two illustrative applications: a small-N case on Internet shutdowns in Africa and a large-N dataset on immigration attitudes in Europe.

One of the key advantages of ARM—particularly in large datasets—lies in its computational efficiency and the ability to uncover a broader set of relevant rules, including those missed by traditional QCA algorithms. We proposed a minimization algorithm for association rules that mimics the Quine–McCluskey approach in csQCA, thus avoiding the need to construct large truth tables.

These findings provide a foundation for researchers to select between csQCA and ARM based on sample size, computational constraints, and research goals. Future research could extend this comparison to Fuzzy-Set QCA and explore the integration of AR-based methods into mixed-method research designs.

Author Contributions: Conceptualization, A.D.L., R.B. and M.d.C.B.; methodology, A.D.L. and M.d.C.B.; software, A.D.L. and R.B.; validation, A.D.L. and R.B.; formal analysis, R.B. and M.d.C.B.; investigation, A.D.L.; data curation, A.D.L.; writing—original draft preparation, A.D.L.; writing—review and editing, R.B. and M.d.C.B.; supervision, R.B. and M.d.C.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study are included in this article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

QCA	Qualitative Comparative Analysis.
csQCA	Crisp-Set QCA.
fsQCA	Fuzzy-Set QCA.
AR	Association rule.
ARM	Association rule mining.

References

1. Ragin, C.C. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*; University of California Press: Oakland, CA, USA, 1987.
2. Pagliarin, S.; La Mendola, S.; Vis, B. The “qualitative” in qualitative comparative analysis (QCA): Research moves, case-intimacy and face-to-face interviews. *Qual. Quant.* **2023**, *57*, 489–507. [[CrossRef](#)]
3. Hanckel, B.; Petticrew, M.; Thomas, J.; Green, J. The use of Qualitative Comparative Analysis (QCA) to address causality in complex systems: A systematic review of research on public health interventions. *BMC Public Health* **2021**, *21*, 877. [[CrossRef](#)] [[PubMed](#)]

4. 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, 25–28 May 1993; Association for Computing Machinery: New York, NY, USA, 1993; SIGMOD '93; pp. 207–216. [\[CrossRef\]](#)
5. Schneider, C.Q.; Wagemann, C. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*; Strategies for Social Inquiry; Cambridge University Press: Cambridge, UK, 2012. [\[CrossRef\]](#)
6. Swiatczak, M.D. Different algorithms, different models. *Qual. Quant.* **2022**, *56*, 1913–1937. [\[CrossRef\]](#)
7. Ragin, C.C.; Sonnett, J. Between Complexity and Parsimony: Limited Diversity, Counterfactual Cases, and Comparative Analysis. In *Vergleichen in der Politikwissenschaft*; VS Verlag für Sozialwissenschaften: Wiesbaden, Germany, 2005; pp. 180–197. [\[CrossRef\]](#)
8. Duşa, A. *QCA with R. A Comprehensive Resource*; Springer International Publishing: Cham, Switzerland, 2019. [\[CrossRef\]](#)
9. Ragin, C.; Rihoux, B. Qualitative comparative analysis (CQA): State of the art and prospects. *Qual. Multi-Method Res.* **2004**, *2*, 3–13. [\[CrossRef\]](#)
10. De Cock, M.; Cornelis, C.; Kerre, E. Elicitation of fuzzy association rules from positive and negative examples. *Fuzzy Sets Syst.* **2005**, *149*, 73–85. [\[CrossRef\]](#)
11. Mueyba, M.; Khan, M.S.; Coenen, F. A Framework for Mining Fuzzy Association Rules from Composite Items. In Proceedings of the New Frontiers in Applied Data Mining, Bangkok, Thailand, 27–30 April 2010; Chawla, S., Washio, T., Minato, S.i., Tsumoto, S., Onoda, T., Yamada, S., Inokuchi, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 62–74.
12. Korhonen-Kurki, K.; Jenniver Sehring, M.B.; Gregorio, M.D. Enabling factors for establishing REDD+ in a context of weak governance. *Clim. Policy* **2014**, *14*, 167–186. [\[CrossRef\]](#)
13. Aversa, P.; Furnari, S.; Haefliger, S. Business model configurations and performance: A qualitative comparative analysis in Formula One racing, 2005–2013. *Ind. Corp. Change* **2015**, *24*, 655–676. [\[CrossRef\]](#)
14. Kumar, S.; Sahoo, S.; Lim, W.M.; Kraus, S.; Bamel, U. Fuzzy-set qualitative comparative analysis (fsQCA) in business and management research: A contemporary overview. *Technol. Forecast. Soc. Change* **2022**, *178*, 121599. [\[CrossRef\]](#)
15. Mangalampalli, A.; Pudi, V. Fuzzy association rule mining algorithm for fast and efficient performance on very large datasets. In Proceedings of the 2009 IEEE International Conference on Fuzzy Systems, Jeju Island, Republic of Korea, 20–24 August 2009; pp. 1163–1168. [\[CrossRef\]](#)
16. Ho, G.; Ip, W.; Wu, C.; Tse, Y. Using a fuzzy association rule mining approach to identify the financial data association. *Expert Syst. Appl.* **2012**, *39*, 9054–9063. [\[CrossRef\]](#)
17. Chien, B.C.; Cheng, M.C. A color image segmentation approach based on fuzzy similarity measure. In Proceedings of the 2002 IEEE World Congress on Computational Intelligence, 2002 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'02. Proceedings (Cat. No.02CH37291), Honolulu, HI, USA, 12–17 May 2002; Volume 1, pp. 449–454. [\[CrossRef\]](#)
18. Tan, P.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*; Always Learning, Pearson: London, UK, 2014.
19. Ragin, C.C.; Shulman, D.; Weinberg, A.S.; Gran, B.K. Complexity, Generality, and Qualitative Comparative Analysis. *Field Methods* **2003**, *15*, 323–340. [\[CrossRef\]](#)
20. Fiss, P.C. A Set-Theoretic Approach to Organizational Configurations. *Acad. Manag. Rev.* **2007**, *32*, 1180–1198. [\[CrossRef\]](#)
21. Woodside, A.G. Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. *J. Bus. Res.* **2013**, *66*, 463–472. [\[CrossRef\]](#)
22. Rihoux, B.; Ragin, C. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2009. [\[CrossRef\]](#)
23. Vis, B. The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analyses. *Sociol. Methods Res.* **2012**, *40*, 168–198. [\[CrossRef\]](#)
24. Seawright, J. Qualitative Comparative Analysis vis-A-vis Regression. *Stud. Comp. Int. Dev.* **2005**, *40*, 3–26. [\[CrossRef\]](#)
25. Buxton, E.K.; Vohra, S.; Guo, Y.; Fogleman, A.; Patel, R. Pediatric population health analysis of southern and central Illinois region: A cross sectional retrospective study using association rule mining and multiple logistic regression. *Comput. Methods Programs Biomed.* **2019**, *178*, 145–153. [\[CrossRef\]](#)
26. Changpetch, P.; Lin, D.K. Model selection for logistic regression via association rules analysis. *J. Stat. Comput. Simul.* **2013**, *83*, 1415–1428. [\[CrossRef\]](#)
27. Zhao, J.; Yan, C. User acceptance of information feed advertising: A hybrid method based on sem and qca. *Future Internet* **2020**, *12*, 209. [\[CrossRef\]](#)
28. Mguirris, I.; Amdouni, H.; Gammoudi, M.M. A new validation method of fuzzy association rules based on the Structural Equation Modeling. In Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, China, 15–17 August 2015; pp. 624–633. [\[CrossRef\]](#)
29. Kim, C.; Costello, F.J.; Lee, K.C. Integrating Qualitative Comparative Analysis and Support Vector Machine Methods to Reduce Passengers' Resistance to Biometric E-Gates for Sustainable Airport Operations. *Sustainability* **2019**, *11*, 5349. [\[CrossRef\]](#)
30. Chiu, A.; Xu, Y. Bayesian Rule Set: A Quantitative Alternative to Qualitative Comparative Analysis. *J. Politics* **2023**, *85*, 280–295. [\[CrossRef\]](#)

31. Álamos Concha, P.; Pattyn, V.; Rihoux, B.; Schalembier, B.; Beach, D.; Cambré, B. Conservative solutions for progress: On solution types when combining QCA with in-depth Process-Tracing. *Qual. Quant.* **2022**, *56*, 1965–1997. [\[CrossRef\]](#)
32. Fu, M. scpQCA: Enhancing mvQCA Applications through Set-Covering-Based QCA Method. *arXiv* **2024**. [\[CrossRef\]](#)
33. Ljungstrom, B.M.; Denk, T.; Sarenmalm, E.K.; Axberg, U. Use of qualitative comparative analysis (QCA) in an explanatory sequential mixed methods design to explore combinations of family factors that could have an impact on the outcome of a parent training program. *Child. Youth Serv. Rev.* **2025**, *170*, 108120. [\[CrossRef\]](#)
34. Merminod, V.; Rowe, F. Identification de configurations par la méthode Fuzzy set Qualitative Comparative Analysis: Illustration par la contribution de la technologie PLM au respect du temps de développement. *Systèmes D'Inf. Manag.* **2018**, *23*, 71–97. [\[CrossRef\]](#)
35. Llopis-Albert, C.; Merigo, J.M.; Xu, Y. Application of Fuzzy Set/Qualitative Comparative Analysis to Public Participation Projects in Support of the EU Water Framework Directive. *Water Environ. Res.* **2018**, *90*, 73–82. [\[CrossRef\]](#)
36. Shalahuddin, M.; Sunindyo, W.; Effendi, M.; Surendro, K. Fuzzy-set qualitative comparative analysis (fsQCA) for validating causal relationships in system dynamics models. *Eng. Rep.* **2024**, *6*, e12855. [\[CrossRef\]](#)
37. Greckhamer, T.; Misangyi, V.F.; Fiss, P.C. The Two QCAs: From a Small-N to a Large-N Set Theoretic Approach. In *Configurational Theory and Methods in Organizational Research*; Research in the Sociology of Organizations; Emerald Group Publishing Limited: Leeds, UK, 2013; Volume 38, pp. 49–75. [\[CrossRef\]](#)
38. Oana, I.E.; Schneider, C.Q.; Thomann, E. *Qualitative Comparative Analysis Using R: A Beginner's Guide*; Methods for Social Inquiry, Cambridge University Press: Cambridge, UK, 2021. [\[CrossRef\]](#)
39. McCluskey, E.J. Minimization of Boolean functions. *Bell Syst. Tech. J.* **1956**, *35*, 1417–1444. [\[CrossRef\]](#)
40. Quine, W.V. The Problem of Simplifying Truth Functions. *Am. Math. Mon.* **1952**, *59*, 521–531. [\[CrossRef\]](#)
41. Quine, W.V. A Way to Simplify Truth Functions. *Am. Math. Mon.* **1955**, *62*, 627–631. [\[CrossRef\]](#)
42. Hegland, M. Algorithms for Association Rules. In *Advanced Lectures on Machine Learning: Machine Learning Summer School 2002 Canberra, Australia, February 11–22. 2002 Revised Lectures*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 226–234. [\[CrossRef\]](#)
43. Hegland, M. The Apriori Algorithm—A Tutorial. In *Mathematics and Computation in Imaging Science and Information Processing*; World Scientific Publishing Co.: Singapore, 2007; pp. 209–262. [\[CrossRef\]](#)
44. Han, J.; Pei, J.; Yin, Y. Mining frequent patterns without candidate generation. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 16–18 May 2000; Association for Computing Machinery: New York, NY, USA, 2000; SIGMOD '00; pp. 1–12. [\[CrossRef\]](#)
45. Zaki, M. Scalable algorithms for association mining. *IEEE Trans. Knowl. Data Eng.* **2000**, *12*, 372–390. [\[CrossRef\]](#)
46. Zaki, M.J.; Hsiao, C.J. CHARM: An Efficient Algorithm for Closed Itemset Mining. In Proceedings of the 2002 SIAM International Conference on Data Mining (SDM), Arlington, VA, USA, 11–13 April 2002; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2002; pp. 457–473. [\[CrossRef\]](#)
47. Das, A.; Ng, W.K.; Woon, Y.K. Rapid association rule mining. In Proceedings of the Tenth International Conference on Information and Knowledge Management, Atlanta, GA, USA, 5–10 November 2001; Association for Computing Machinery: New York, NY, USA, 2001; CIKM '01; pp. 474–481. [\[CrossRef\]](#)
48. Savasere, A.; Navathe, S.; Omiecinski, E. Mining for Strong Negative Associations in a Large Database of Customer Transactions. In Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE), Brisbane, Australia, 8–12 April 2013; IEEE Computer Society: Los Alamitos, CA, USA, 1998; p. 494. [\[CrossRef\]](#)
49. Wu, X.; Zhang, C.; Zhang, S. Efficient Mining of Both Positive and Negative Association Rules. *ACM Trans. Inf. Syst.* **2004**, *22*, 381–405. [\[CrossRef\]](#)
50. Khedr, A.; Ramadan, R.; Abdel-Mageid, S. Qmr: Quine-mccluskey for rule minimization in rule-based systems. *Int. J. Intell. Comput. Inf. Sci.* **2012**, preprint.
51. Freyburg, T.; Garbe, L. Authoritarian Practices in the Digital Age! Blocking the Bottleneck: Internet Shutdowns and Ownership at Election Times in Sub-Saharan Africa. *Int. J. Commun.* **2018**, *12*, 21.
52. Cornelis, C.; Yan, P.; Zhang, X.; Chen, G. Mining Positive and Negative Association Rules from Large Databases. In Proceedings of the 2006 IEEE Conference on Cybernetics and Intelligent Systems, Taipei, Taiwan, 8–11 October 2006; pp. 1–6. [\[CrossRef\]](#)
53. Emmenegger, P.; Schraff, D.; Walter, A. QCA, the Truth Table Analysis and Large-N Survey Data: The Benefits of Calibration and the Importance of Robustness Tests. In Proceedings of the 2nd International QCA Expert Workshop, Zurich, Switzerland, 5–7 November 2014.
54. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
55. Hahsler, M.; Gruen, B.; Hornik, K. Arules—A Computational Environment for Mining Association Rules and Frequent Item Sets. *J. Stat. Softw.* **2005**, *14*, 1–25. [\[CrossRef\]](#)
56. Hahsler, M.; Chelluboina, S.; Hornik, K.; Buchta, C. The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Datasets. *J. Mach. Learn. Res.* **2011**, *12*, 1977–1981.

57. Mendel, J.M.; Korjani, M.M. Theoretical aspects of Fuzzy Set Qualitative Comparative Analysis (fsQCA). *Inf. Sci.* **2013**, *237*, 137–161. [[CrossRef](#)]
58. Agrawal, A.; Knoeber, C.R. Firm Performance and Mechanisms to Control Agency Problems between Managers and Shareholders. *J. Financ. Quant. Anal.* **1996**, *31*, 377–397. [[CrossRef](#)]
59. Ragin, C.C. *Redesigning Social Inquiry*; University of Chicago Press: Chicago, IL, USA, 2008.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.