Ingénierie des Systèmes d'Information
Vol. 31, No. 10, October, 2026, pp. 3273-3281
Journal homepage: http://iieta.org/journals/isi

Modeling Conditional Dependencies in Big Data Analytics Adoption among Large Taxpayers in Albania: A TOE-Based Bayesian Network Analysis

Megi Marku

Department of Mathematics and Informatics, Faculty of Economy and Agribusiness, Agricultural University of Tirana, Tirana 1025, Albania

Corresponding Author Email: mmarku@ubt.edu.al

Copyright: ©2026 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/isi.311003

Received: 30 June 2026
Revised: 8 September 2026
Accepted: 15 September 2026
Available online: 31 October 2026

Keywords:
big data analytics adoption, Bayesian networks, Technology–Organization–Environment framework, conditional dependencies, organizational technology adoption, emerging economies, Albania

ABSTRACT

Big data analytics (BDA) adoption is shaped by interrelated technological, organizational, and environmental conditions. However, conventional adoption studies often estimate prespecified relationships without revealing the wider dependency structure among these factors. This study examines conditional dependencies associated with BDA adoption by integrating the Technology–Organization–Environment (TOE) framework with Bayesian network analysis. Survey data were collected from 121 firms listed in Albania's Large Taxpayers Directorate register, comprising 51 adopters and 70 non-adopters. Nine TOE determinants and a binary adoption-status variable were analyzed using score-based hill-climbing structure learning with the Bayesian information criterion. TOE scores missing by design for non-adopters were estimated using the expectation-maximization (EM) algorithm under a missing-at-random working assumption. Network stability was assessed through 500 bootstrap resamples, and the Markov blanket was used to identify determinants most directly connected to adoption. The resulting network places BDA adoption within an interconnected probabilistic structure. Its Markov blanket comprises complexity, risk and insecurity, top management support, and government support. Top management support and complexity show the strongest adoption-related dependencies, with bootstrap strengths of 0.904 and 0.858, respectively. Relative advantage is connected to compatibility and competitive pressure in the wider network. In 10-fold cross-validation, the Bayesian network achieved an area under the receiver operating characteristic curve (AUC) of 0.867 (95% confidence interval (CI): 0.796–0.939), compared with 0.806 (95% CI: 0.697–0.915) for logistic regression; overlapping intervals do not establish superior predictive performance. Given the modest sample and imputation assumption, the network should be interpreted as exploratory and context-specific.

1. INTRODUCTION

The growing sophistication of digital transformation and data-driven organizational environments has elevated the strategic importance of big data analytics (BDA) in both established and emerging economies. Organizations are increasingly reliant on analytics technologies to improve operational efficiency, strengthen decision-making, maximize resource allocation, and ensure long-term competitiveness. As the volume, velocity, and variety of organizational data increase, firms encounter heightened demand to transform raw data into meaningful insights that support strategic and operational objectives [1, 2].

Consequently, BDA is now regarded not merely as a technological innovation but as an essential organizational capability required for sustaining competitiveness in dynamic and uncertain business environments. Despite the recognized benefits of analytics technology, the adoption of BDA remains inconsistent, particularly in emerging economies characterized by infrastructural inadequacies, institutional volatility, financial constraints, and insufficient digital readiness [3].

Prior studies indicate that successful analytics implementation depends on technological infrastructure, organizational capabilities, managerial support, and external environmental support systems [4, 5]. The adoption of BDA is a complex process shaped by interconnected technological, organizational, and environmental factors. The Technology–Organization–Environment (TOE) framework, developed by Tornatzky and Fleischer [6], has become a leading theoretical model for examining organizational technology adoption.

The TOE framework asserts that adoption decisions are influenced by technological factors such as relative advantage, compatibility, complexity, and perceived risk; organizational elements including executive support and readiness; and environmental factors like competitive pressure, governmental support, and external assistance. Although the TOE framework has become one of the dominant theoretical lenses for explaining organizational technology adoption, empirical studies on BDA have been dominated by linear explanatory techniques, particularly logistic regression, covariance-based structural equation modeling (CB-SEM), and partial least squares structural equation modeling (PLS-SEM) [7].

These approaches have substantially advanced understanding of the determinants of BDA adoption by estimating direct, indirect, and mediating relationships among technological, organizational, and environmental constructs. However, they generally require the causal structure to be specified a priori and primarily evaluate predefined relationships among latent variables. As a result, they offer limited insight into the probabilistic dependency structure that may simultaneously emerge among TOE determinants during organizational technology adoption [4].

Previous TOE-based BDA studies consistently highlight the importance of these factors in clarifying organizational analytics adoption behavior [7]. Existing TOE literature further suggests that adoption determinants rarely operate independently. Greater technological complexity may reduce organizational readiness because organizations require additional technical expertise, financial resources, and implementation capabilities.

Likewise, higher perceived technological risk may weaken organizational readiness by increasing uncertainty surrounding implementation outcomes. Top management support may strengthen compatibility by facilitating organizational change, resource allocation, and strategic alignment, whereas perceived relative advantage may influence the way organizations interpret competitive pressure by increasing managerial awareness of strategic opportunities.

Collectively, these findings suggest that TOE determinants form an interconnected organizational system, providing a strong theoretical rationale for modeling their conditional probabilistic relationships using Bayesian networks rather than considering them as isolated predictors. Within the TOE framework, technological complexity, compatibility, organizational readiness, top management support, and competitive pressure frequently function through sequential and conditional interactions rather than independent impacts [8].

Nevertheless, most contemporary research primarily uses regression-based methodologies or SEM techniques that treat adoption drivers as substantially independent predictors. For instance, Maroufkhani et al. [4] applied PLS-SEM to examine BDA adoption determinants among small and medium-sized enterprises (SMEs), while Lutfi et al. [3] and Aldossari et al. [5] similarly relied on regression-based or survey-based statistical techniques to model adoption as a function of largely independent predictors.

Bayesian networks have been applied extensively across diverse domains, including military decision support [9] and healthcare [10], demonstrating their broad methodological versatility. By contrast, applications of Bayesian networks within technology adoption research remain comparatively rare and are largely confined to adjacent domains such as software acceptance [11], general information and communication technology (ICT) adoption [12], and online review adoption [13], rather than BDA adoption specifically within the TOE framework. This pattern suggests that, despite growing recognition of the interconnected nature of adoption determinants, probabilistic graphical modeling remains underutilized in this specific research context.

More recently, configurational approaches such as fuzzy-set qualitative comparative analysis (fsQCA) have gained increasing attention in technology adoption research because they acknowledge that multiple combinations of technological, organizational, and environmental conditions may equally lead to successful adoption [7].

Unlike regression or SEM, fsQCA embraces causal asymmetry and equifinality by identifying sufficient and necessary configurations of conditions. Nevertheless, although configurational approaches improve understanding of alternative adoption pathways, they do not explicitly model conditional probabilistic dependencies among determinants or continuously update adoption probabilities when new evidence becomes available. Consequently, important interrelationships among TOE factors may remain hidden.

Although these methodologies provide valuable explanatory insights, they may overly simplify the inherently interconnected and uncertainty-driven facets of organizational technology adoption. The complexity of technology may impact organizational readiness, managerial support may depend on technology compatibility, and environmental factors may indirectly affect adoption decisions through institutional support systems.

Accordingly, BDA adoption can be examined as part of a probabilistic system characterized by conditional dependencies and nonlinear interrelations among determinants, rather than as an assemblage of distinct linear correlations. Bayesian networks provide a methodological framework suited to modeling complex conditional dependencies. These probabilistic graphical models complement rather than replace existing analytical approaches.

While regression models estimate average effects, SEM focuses on latent-variable relationships, and fsQCA identifies alternative causal configurations, Bayesian networks offer a distinct probabilistic alternative. Within technology adoption research specifically, such networks explicitly model conditional probabilistic dependencies among variables [11-13]. Their graphical representation enables the identification of interconnected pathways through which technological, organizational, and environmental determinants are jointly associated with adoption decisions under uncertainty.

Furthermore, Bayesian networks naturally accommodate incomplete information and support probabilistic updating as new evidence becomes available, making them particularly suitable for complex organizational decision environments [12]. Their directed acyclic graph (DAG) representation enables conditional relationships among variables to be examined within a unified probabilistic structure.

Unlike conventional statistical methods, Bayesian approaches facilitate the modeling of uncertainty, indirect effects, and interconnected probabilistic pathways within complex organizational systems. Bayesian networks have increasingly been employed in healthcare, decision support systems, risk analysis, knowledge management, and technology adoption studies because of their ability to model dynamic interactions in uncertain situations [11].

Bayesian networks have been used to model information systems marked by intricate interdependencies across technological, organizational, and environmental factors. Prior research demonstrates that Bayesian networks can identify probabilistic associations among demographic, social, and behavioral factors associated with technology adoption in emerging and digitally constrained contexts [12].

Moreover, Bayesian methodologies have been increasingly used to assess technological acceptance, message adoption, and perceived credibility in digital platforms, underscoring their adaptability for assessing multifaceted adoption behavior amid uncertainty [13].

Prior research has used Bayesian approaches to examine

complex linkages among socioeconomic, organizational, and behavioral factors impacting technology adoption processes. Notwithstanding their methodological benefits, Bayesian networks are markedly underrepresented in research concerning organizational technology adoption. Prior studies indicate that the effective utilization of Bayesian networks is sometimes obstructed by methodological complexities, a lack of standardized development protocols, and inadequate integration into corporate decision-making processes [12].

Empirical studies examining BDA adoption through Bayesian probability frameworks remain scarce, particularly in emerging economies. This study addresses the research gap by integrating Bayesian network modeling with the TOE framework to analyze the probabilistic interdependencies among factors influencing BDA adoption among large taxpayers in Albania.

Despite the considerable body of research examining BDA adoption through the TOE framework, several important theoretical gaps remain.

First, existing studies primarily estimate direct relationships between TOE determinants and adoption outcomes, paying comparatively little attention to the dependency structure among the determinants themselves. Second, although configurational approaches have broadened the understanding of technology adoption by recognizing multiple adoption pathways, they remain less suited for representing probabilistic interactions under uncertainty. Third, Bayesian networks remain rarely employed in organizational technology adoption research despite their demonstrated ability to model interconnected systems characterized by uncertainty, indirect effects, and conditional dependencies [11].

Addressing these limitations may provide a richer theoretical understanding of organizational technology adoption by conceptualizing TOE determinants as an interconnected probabilistic structure rather than as isolated explanatory variables.

This study addresses the following research questions (RQs): (RQ1) What is the conditional dependency structure among TOE determinants and BDA adoption, beyond their individual direct associations with adoption? (RQ2) Which TOE determinants form the immediate probabilistic neighborhood of BDA adoption—that is, its Markov blanket—once dependencies among all determinants are simultaneously accounted for? These questions are not directly addressed by conventional regression or prespecified SEM approaches, which primarily evaluate hypothesized relationships rather than learning the joint probabilistic dependency structure among all variables. A Bayesian network is therefore particularly suitable for addressing the dependency structure and Markov blanket examined in this study.

Albania provides a particularly relevant empirical context for investigating organizational BDA adoption. Although the country has made substantial progress in digital transformation in recent years, the adoption of advanced analytical technologies remains relatively limited compared with more digitally mature European economies. Organizations in emerging economies frequently experience constraints related to technological infrastructure, analytical capabilities, organizational readiness, and institutional support [3, 14].

Consequently, Albania provides an appropriate setting for examining how technological, organizational, and environmental conditions jointly relate to BDA adoption under the constraints typically observed in emerging economies. These contextual characteristics also contribute to extending the geographical diversity of information systems research, where evidence from Southeast European economies remains comparatively scarce.

The study seeks to deepen the understanding of organizational analytics adoption under uncertainty by characterizing TOE factors as interconnected probabilistic pathways rather than as separate predictors.

This study contributes to the current literature in several dimensions. First, this research contributes to TOE-based BDA adoption studies by using a systems-oriented probabilistic approach that captures nonlinear conditional dependencies among adoption determinants, while acknowledging the exploratory nature of the resulting structure. Second, it contributes methodologically by demonstrating the applicability of Bayesian networks to modeling organizational technology adoption under uncertainty. Third, the study presents empirical evidence from an emerging economy where challenges related to digital transformation and institutional constraints consistently hinder the adoption of organizational analytics. Finally, the study responds to recent calls for methodological diversification in organizational technology adoption research by demonstrating the applicability of Bayesian probabilistic modeling within the TOE framework.

The findings aim to provide theoretical and practical insights for managers and policymakers seeking to improve BDA implementation across complex organizational ecosystems.

2. METHODOLOGY

This study employed a quantitative and probabilistic modeling approach to analyze the interconnections among the TOE determinants influencing BDA adoption. The empirical study employed survey data collected from 121 large taxpayers in Albania. Of 250 questionnaires distributed to firms in the Large Taxpayers Directorate register (N = 840), 125 were returned (questionnaire return rate = 125/250 = 50.0%), and 121 valid responses were retained (valid response rate = 121/250 = 48.4%), of which 51 firms (42.1%) were BDA adopters and 70 (57.9%) were non-adopters.

Large taxpayers are generally characterized by greater organizational complexity, higher digitalization requirements, and stronger investment capacity, making them particularly suitable for investigating advanced analytics adoption.

The technological dimension included relative advantage, compatibility, complexity, and risk and insecurity. The organizational dimension included top management support and organizational readiness. The environmental dimension included competitive pressure, external support, and government support.

In addition to the binary adoption-status variable, the questionnaire included four items assessing the intensity and centrality of BDA use within daily business operations, capturing: (i) the intensity of BDA use in supporting operations, (ii) the degree of BDA's influence on operations, (iii) the extent of operational dependence on BDA, and (iv) the use of BDA in decision-making. Together, they reflect the extent of BDA deployment and organizational reliance on analytics in daily operations, consistent with the

conceptualization of adoption as a matter of degree.

However, for the Bayesian network analysis, the Adoption node was operationalized using the observed adopter/non-adopter status of each firm, thereby providing a directly observed binary target — a distinction elaborated below. The construct measurement items used a five-point Likert scale, where 1 represented "strongly disagree" and 5 signified "strongly agree." Composite construct scores were calculated by averaging the pertinent measurement items for each TOE determinant.

The questionnaire employed a skip-logic design with two respondent pathways: respondents who reported using BDA completed the TOE-specific items, whereas non-adopters were routed past these items — as they presuppose direct experience with the technology — and proceeded directly to items assessing general perceptions of BDA usefulness and ease of use, along with firm performance.

As a result, TOE construct scores were structurally missing for non-adopting firms, whereas the adoption-status variable, perception items, and performance items were collected from all 121 respondents. Because the Bayesian network in this study relies on the TOE constructs, missing values for non-adopters were addressed using the expectation-maximization (EM) algorithm described below.

The EM algorithm operated under a missing-at-random working assumption, preserving the full analytical sample of n = 121. As EM imputation for a structurally missing subgroup rests on assumptions the design cannot fully verify, the resulting network is treated as exploratory and hypothesis-generating rather than confirmatory.

Descriptive statistics and preliminary data screening methods were conducted to ensure data quality and consistency. For probabilistic network modeling, the continuous composite scores were discretized into two categories (Low/High) at the median.

A three-level scheme was initially tested but rejected. The imputed composites contained numerous tied values, which made quantile binning infeasible for several constructs, and equal-interval binning produced sparse cells (as few as two firms) that destabilize conditional-probability estimation in a modest sample.

The two-level scheme yielded workable cell frequencies across all constructs. The discretization approach was implemented because discrete Bayesian network estimation necessitates categorical representations of variables to enable probabilistic dependency learning and conditional probability estimation. Comparable discretization methods have been employed in previous Bayesian network research investigating organizational and technological systems under uncertainty.

While necessary to avoid sparse cells given the sample size, discretizing continuous composite scores into binary Low/High categories at the median entails some loss of within-category variability and may classify firms near the median boundary somewhat arbitrarily. This represents an additional reason, alongside the observations-to-node ratio, to treat the resulting network structure as exploratory rather than confirmatory.

The adoption node was operationalized as the observed adopter/non-adopter status of each firm (51 adopters, 70 non-adopters), avoiding an arbitrary threshold on a self-reported scale and yielding a binary target. The final dataset utilized in the Bayesian analysis comprised exclusively categorical variables, as required for discrete Bayesian network estimation.

The study employed Bayesian network modeling to determine probabilistic associations and structural linkages between TOE determinants and BDA adoption. Bayesian networks are graphical probabilistic models that depict conditional interdependence among variables through DAGs. Unlike standard regression approaches that primarily estimate prespecified associations between predictors and an outcome, Bayesian networks enable the joint representation of conditional probabilistic dependencies among variables.

The Bayesian network structure was determined using the hill-climbing (HC) score-based learning method, as executed in the bnlearn package in R. The HC approach was chosen for its appropriateness in exploratory probabilistic structure learning with multiple potentially interdependent variables.

Prior investigations utilizing Bayesian networks have predominantly implemented score-based search techniques alongside Bayesian information criterion (BIC) optimization to identify parsimonious dependency structures while controlling model complexity. The HC algorithm iteratively searches for a network structure that optimizes the BIC score. The resulting network represents the estimated dependency structure between TOE determinants and BDA adoption.

The Markov blanket of the adoption node was further estimated to evaluate the local probabilistic environment for BDA adoption. The Markov blanket defines the set of variables that contains all relevant probabilistic information directly associated with the target node.

This method enables the identification of the TOE determinants most closely linked to BDA adoption within the network structure. Bootstrap resampling with 500 replicates was used to assess the stability of the learned Bayesian network structure with the boot.strength() function of the bnlearn package.

Bootstrap resampling was used to evaluate the stability of the learned probabilistic dependencies by measuring the frequency with which individual arcs were recovered across resampled datasets. A threshold of 0.50 was applied, and arcs exceeding this threshold were retained in the bootstrap-averaged Bayesian network.

Network visualization was used to illustrate the stable probabilistic dependencies between TOE determinants and BDA adoption. All analyses were conducted with R statistical software. While Bayesian networks facilitate probabilistic reasoning and interdependency analysis, the established linkages should not be construed as unequivocally causal associations.

The learned network structure predominantly represented conditional probabilistic dependencies based on the observed dataset. The combination of HC optimization, BIC-based structural assessment, Markov blanket analysis, and bootstrap-averaged Bayesian networks facilitated the development of an exploratory probabilistic framework for analyzing the interrelated TOE determinants affecting BDA adoption.

3. RESULTS

This section presents the empirical findings in four stages. It first reports the reliability of the measurement constructs, then the structure of the learned Bayesian network and the Markov blanket of BDA adoption, followed by the bootstrap stability and threshold-sensitivity analyses that assess the robustness of the recovered structure.

Finally, it reports the predictive performance of the network

against a logistic-regression baseline. Together, these results characterize BDA adoption as a conditional, interconnected structure among technological, organizational, and environmental determinants rather than a set of independent effects.

Prior to network estimation, the internal consistency of the nine TOE constructs and the four-item BDA adoption intensity scale was assessed. All constructs exceeded the 0.70 threshold (α = 0.845–0.962), and the Kaiser–Meyer–Olkin (KMO) measure (0.795) with a significant Bartlett's test ($p < 0.001$) indicated sampling adequacy (Table 1). Convergent validity was assessed using average variance extracted (AVE), which ranged from 0.721 to 0.880 across all nine TOE constructs, exceeding the minimum recommended threshold of 0.50.

Discriminant validity was confirmed using both the Fornell–Larcker criterion, whereby the square root of each construct's AVE exceeded its correlations with all other constructs, and the heterotrait–monotrait ratio (HTMT), with all values falling below the recommended threshold of 0.85.

These results, derived from the broader validated measurement model underlying this dataset, confirmed that the TOE constructs are both theoretically and empirically distinct.

Table 1. Construct reliability

Construct	Items	Cronbach's α
Relative advantage (RA)	6	0.962
Compatibility (CO)	3	0.886
Complexity (C)	3	0.900
Risk and insecurity (RI)	4	0.955
Top management support (M)	4	0.909
Organizational readiness (OR)	3	0.887
Competitive pressure (CP)	3	0.845
External support (ES)	3	0.886
Government support (G)	3	0.854
BDA adoption intensity (A)	4	0.859

Note: BDA = Big data analytics. BDA adoption intensity (A) is a four-item scale capturing the intensity and centrality of BDA use in daily operations; it is reported here for reliability assessment only and is distinct from the binary Adoption node (observed adopter/non-adopter status) used in the Bayesian network. Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy = 0.795; Bartlett's test of sphericity, $p < 0.001$.

Table 2. Bayesian network structure identified through the hill-climbing (HC) algorithm

From	To
C	RI
C	OR
RI	OR
C	Adoption
RI	Adoption
Adoption	M
RI	G
Adoption	G
C	CO
M	CO
G	ES
CO	RA
M	RA
RA	CP
ES	CP

Note: RA = relative advantage; CO = compatibility; C = complexity; RI = risk and insecurity; M = top management support; OR = organizational readiness; CP = competitive pressure; ES = external support; G = government support; Adoption = big data analytics (BDA) adoption.

Table 3. Markov blanket of big data analytics (BDA) adoption

Target Variable	Markov Blanket Variables
Adoption	Complexity (C), risk and insecurity (RI), top management support (M), government support (G)

Table 2 presents the Bayesian network structure initially identified through the HC algorithm. The results indicate that BDA Adoption is not conditioned by a single antecedent but sits within a cluster of directly connected determinants. Complexity and risk and insecurity are directly linked to Adoption, which is in turn connected to top management support and government support — together forming its immediate probabilistic neighborhood, consistent with the Markov blanket reported in Table 3.

It is important to clarify that the specific arc orientations learned by the structure-learning algorithm reflect statistical dependencies identified within a Markov equivalence class of graphs, rather than confirmed causal ordering. Bayesian network structure learning from cross-sectional observational data typically cannot distinguish between statistically equivalent orientations without additional temporal or interventional information.

Consistent with established TOE theory, top management support and complexity are conceptually treated as antecedent conditions influencing adoption; the learned arc direction should therefore be interpreted as reflecting the strength and existence of a probabilistic dependency, not as evidence that adoption causally precedes these factors.

This configuration indicates that adoption is directly associated within the network with perceived implementation difficulty (complexity), perceived uncertainty (risk and insecurity), managerial sponsorship (top management support), and institutional context (government support).

Around this core, the network reveals an interconnected dependency structure: Complexity is further connected to organizational readiness and, together with top management support, is linked to compatibility, which in turn is associated with relative advantage. Relative advantage and external support jointly connect to competitive pressure, while government support links to external support.

Taken together, these results indicate that the TOE determinants function as a cohesive and interrelated probabilistic structure rather than as independent predictors.

Table 3 illustrates the Markov blanket of the BDA adoption node—the set of variables that encompasses all probabilistic information directly related to the target variable, comprising complexity (C), risk and insecurity (RI), top management support (M), and government support (G). These four determinants carry the information most directly associated with adoption within the network; conditional on them, the remaining TOE variables add little further probabilistic information.

The prominence of top management support and complexity indicates that, within the estimated network, adoption has its most stable connections with organizational sponsorship and perceived implementation difficulty, consistent with the roles these factors are typically assigned in the TOE literature. As with all relationships in this network, this association reflects conditional dependence rather than a confirmed causal mechanism.

Table 4 reports bootstrap arc strengths together with 95% confidence intervals, calculated using the standard binomial-

proportion approximation given that each strength value represents the proportion of 500 bootstrap resamples in which the corresponding arc was recovered.

The averaged network retained 14 stable arcs. The strongest adoption-related dependency linked top management support to adoption (0.904), followed by complexity (0.858), government support (0.738), and risk and insecurity (0.574).

Bootstrap directionality values ranged from 0.51 to 0.73 (Table 4), consistent with interpreting these arcs as conditional dependencies rather than directed causal effects.

Table 4. Stable arcs in the bootstrap-averaged network (500 resamples, strength ≥ 0.50)

From	To	Strength	95% CI	Direction
RA	CP	0.994	0.987–1.000	0.538
C	OR	0.794	0.759–0.829	0.725
C	RI	0.736	0.697–0.775	0.557
G	ES	0.720	0.681–0.759	0.636
RI	OR	0.688	0.647–0.729	0.658
M	RA	0.648	0.606–0.690	0.693
M	CO	0.608	0.565–0.651	0.571
CO	RA	0.598	0.555–0.641	0.532
ES	CP	0.578	0.535–0.621	0.588
RI	G	0.552	0.508–0.596	0.639
Adoption	M	0.904	0.878–0.930	0.627
Adoption	C	0.858	0.827–0.889	0.505
Adoption	G	0.738	0.699–0.777	0.688
Adoption	RI	0.574	0.531–0.617	0.564

Note: CI = confidence interval; RA = relative advantage; CO = compatibility; C = complexity; RI = risk and insecurity; M = top management support; OR = organizational readiness; CP = competitive pressure; ES = external support; G = government support; Adoption = Big data analytics (BDA) adoption. Strength = proportion of the 500 bootstrap resamples in which an arc between the two nodes was present, irrespective of direction; Direction = conditional proportion of those resamples in which the arc had the displayed orientation, given that the arc was present.

A sensitivity analysis across thresholds (0.40–0.70) confirmed that the core structure — particularly the adoption dependencies on top management support and complexity — persisted at stricter cut-offs, supporting the robustness of the recovered network (Table 5).

Table 5. Threshold sensitivity of the averaged network

Threshold	Arcs Retained
0.40	17
0.50	14
0.60	10
0.70	7

Table 6. Predictive performance (10-fold cross-validation)

Model	AUC	95% CI	Accuracy	Sensitivity	Specificity
Bayesian network	0.867	0.796–0.939	0.826	0.745	0.886
Logistic regression (baseline)	0.806	0.697–0.915	—	0.804	1.000

Note: AUC = area under the receiver operating characteristic curve; CI = confidence interval.

To assess predictive relevance, the network's classification of the adoption node was evaluated via 10-fold cross-validation, yielding an area under the receiver operating characteristic curve (AUC) of 0.867 (95% confidence interval (CI) 0.796–0.939) and 82.6% accuracy. A logistic-regression baseline using the same predictors achieved a comparable AUC of 0.806 with overlapping confidence intervals (Table 6). Given this overlap, no statistically meaningful difference in discriminative performance between the two models can be claimed.

The Bayesian network yielded a numerically higher AUC, although the overlapping intervals do not establish a difference in predictive performance; its additional value lies in recovering the conditional-dependency structure among the TOE determinants, a form of insight a single-equation model cannot provide.

This structural contribution, rather than raw predictive superiority, is the methodological value of the probabilistic approach.

In Figure 1, each node represents one of the nine TOE determinants or the BDA Adoption variable, and an edge indicates a conditional dependency retained through bootstrap-averaged structure learning (Table 4). Edges are displayed without directional arrows, since the bootstrap-averaged directionality values ranged only from 0.51 to 0.73 — supporting the existence of these dependencies more than their precise orientation; readers are referred to Table 4 for each arc's estimated direction.

Figure 1 shows that the probabilistic linkages among the TOE determinants form a cohesive, interconnected structure rather than discrete associations. BDA adoption occupies a central position, conditionally associated with top management support, complexity, government support, and risk and insecurity — its Markov blanket. The technological determinants remain influential within the wider network: relative advantage connects to competitive pressure and compatibility, while complexity links to organizational readiness and risk and insecurity, with top management support and complexity forming the most stable adoption-related dependencies. Together, these patterns indicate that BDA adoption is embedded within an interconnected probabilistic structure rather than a single linear pathway.

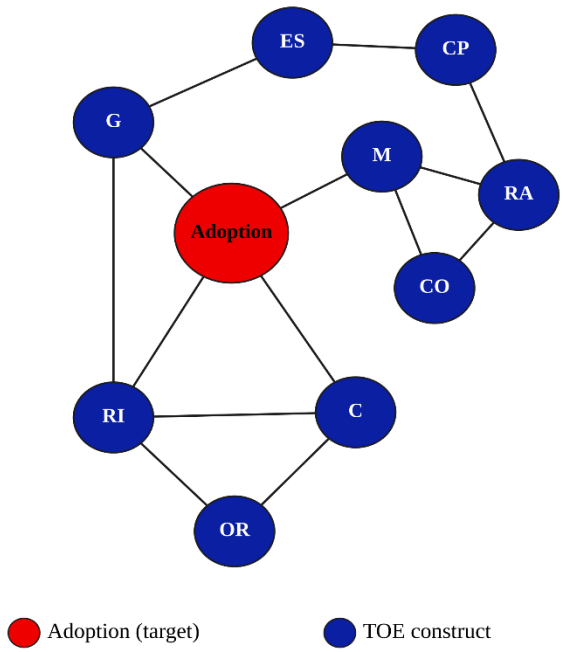


Figure 1. Network structure of Technology–Organization–Environment (TOE) determinants and big data analytics (BDA) adoption

4. DISCUSSION

This study analyzed the probabilistic interconnections among TOE determinants influencing BDA adoption through a Bayesian network approach. Unlike traditional TOE-based research that primarily examines prespecified relationships among adoption determinants through regression or SEM methods, the findings suggest that BDA adoption is embedded within a structure of conditional dependencies among technological, organizational, and environmental factors, rather than being represented through isolated, prespecified relationships.

Taken together, these findings position BDA adoption within an interconnected probabilistic structure — anchored by complexity, risk and insecurity, top management support, and government support — rather than as the outcome of independent, isolated predictors. The Markov blanket analysis identifies these four determinants as the ones most directly associated with adoption.

Top management support shows the strongest adoption-related dependency, underscoring managerial sponsorship as a central factor in BDA adoption, while complexity and risk and insecurity represent implementation- and uncertainty-related factors directly associated with adoption. This result corresponds with prior research suggesting that technological complexity constitutes a major barrier to analytics implementation, particularly in developing countries characterized by inadequate digital maturity and limited technology expertise [3, 15].

The framework suggests that organizations view complexity not merely as a technical issue but as a broader challenge intertwined with organizational readiness and perceived risk, resonating with research conceptualizing BDA capability as an organizational competence built on interrelated tangible, human, and intangible resources rather than technology alone [1]. The remaining technological determinants connect to this core indirectly: relative advantage is associated with compatibility and competitive pressure rather than directly with adoption, indicating that perceived technological benefits operate within a broader network of relationships rather than in isolation, consistent with adoption operating through interconnected conditional dependencies rather than independent technology perceptions.

The bootstrap stability analysis corroborates this robustness, with the strongest, most stable dependencies linking adoption to top management support and complexity, and complexity to risk and insecurity — indicating that these determinants form the most resilient component of the BDA adoption ecosystem, consistent with studies showing that technological and organizational considerations frequently dominate analytics adoption in developing countries facing infrastructure and implementation challenges [5].

The organizational dimension operates in close connection with the adoption core. Top management support is directly linked to adoption within the Markov blanket, while also connecting to compatibility and relative advantage, indicating that managerial commitment is directly associated with adoption and is also connected to surrounding technological perceptions. This is consistent with evidence that the strategic alignment of analytics capability with organizational direction is central to realizing performance gains from BDA [16]. Organizational readiness is probabilistically associated with complexity and risk and insecurity, indicating that perceptions of internal readiness are closely associated with implementation challenges and perceived technological risk.

The environmental dimension demonstrates interconnected rather than isolated effects: Government support forms part of the adoption structure and connects to external support, which in turn links to competitive pressure. This suggests that environmental determinants operate largely through interconnected institutional support mechanisms rather than direct, independent adoption pathways — aligning with prior studies indicating that environmental determinants generally exert context-dependent effects in emerging economies characterized by developing institutional digital ecosystems [7].

Methodologically, the study contributes to the BDA and TOE literature [17] by demonstrating the value of Bayesian networks for modeling organizational technology adoption under uncertainty, an area where prior TOE-based research has predominantly relied on linear methodologies estimating prespecified relationships among adoption determinants.

Rather than arguing that Bayesian networks outperform regression, SEM, or configurational techniques, the findings indicate that they offer complementary methodological value by uncovering dependency structures — and a systemic, probabilistic view of adoption — that remain difficult to observe using conventional approaches. The integration of bootstrap averaging and Markov blanket analysis further strengthens the stability assessment and interpretability of this framework, positioning Bayesian networks as an additional analytical tool for enriching future TOE-based technology adoption research.

Several limitations frame the exploratory character of this study and mean the recovered network should be regarded as context-specific and provisional. First, the sample is modest relative to the number of nodes (10 nodes, n = 121; ratio ≈ 12:1), which the bootstrap arc-strength assessment and threshold-sensitivity evaluation alleviate but do not eliminate [18], limiting confidence in the precise arcs recovered. Second, missing values from the skip-logic design were estimated using EM, which relies on assumptions the study design cannot fully verify and which qualify relationships involving TOE constructs among non-adopters [19]. Third, the bootstrap directionality values (0.51–0.73) reinforce the exploratory, non-causal framing adopted throughout. Fourth, constructs were self-reported at a single point in time, precluding confirmation of whether these dependencies remain stable over time.

From a practical standpoint, managers and governments seeking to enhance BDA adoption may consider prioritizing managerial sponsorship, implementation simplicity, and strategies for mitigating uncertainty. Given that top management support and complexity are most directly associated with adoption, businesses should secure managerial commitment for analytics initiatives while reducing perceived implementation complexity through better integration of analytics technologies with existing operational systems. Evidence from other settings likewise links BDA adoption to organizational performance gains when supported by complementary knowledge capabilities [20].

The network structure offers a concrete example of how such prioritization could be sequenced: Complexity is linked directly to adoption and is also among its two strongest secondary connections, linking to organizational readiness and risk and insecurity. This suggests that interventions aimed at reducing perceived implementation complexity may also be relevant to organizational readiness and perceived risk,

although not causally, illustrating the distinctive practical value of the Bayesian network's dependency structure for identifying clusters of factors that may warrant joint managerial attention. Because government support and risk and insecurity also form part of the adoption structure, institutional support mechanisms and uncertainty-mitigation strategies remain important complementary levers.

5. CONCLUSIONS

This study examined BDA adoption among large taxpayers in Albania by integrating the TOE framework with Bayesian network analysis. The results indicate that BDA adoption is embedded within a complex probabilistic structure in which technological, organizational, and environmental determinants are interconnected. Its Markov blanket comprises complexity, risk and insecurity, top management support, and government support, with top management support and complexity forming the most stable adoption-related connections, while the remaining determinants operate largely through indirect conditional dependencies.

Given the modest sample, the EM-based estimation of structurally missing TOE scores, and the cross-sectional design, the recovered network should be regarded as exploratory and context-specific. Future work with larger, longitudinal, multi-source data would allow this exploratory structure to be tested confirmatorily; until then, applying it to inform decision-making in other organizational or national contexts without independent validation would be premature.

By combining Bayesian network modeling with the TOE framework, the research provides a systems-oriented probabilistic methodology that contributes to organizational technology adoption theory and offers practical insights for digital transformation strategies in Albania, with potential relevance to other emerging-economy settings, subject to independent validation.

REFERENCES

[1] Gupta, M., George, J.F. (2016). Toward the development of a big data analytics capability. Information & Management, 53(8): 1049-1064. https://doi.org/10.1016/j.im.2016.07.004

[2] Wamba, S.F., Gunasekaran, A., Akter, S., Ren, S.J., Dubey, R., Childe, S.J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 70: 356-365. https://doi.org/10.1016/j.jbusres.2016.08.009

[3] Lutfi, A., Alsyouf, A., Almaiah, M.A., Alrawad, M., Abdo, A.A.K., Al-Khasawneh, A.L., Ibrahim, N., Saad, M. (2022). Factors influencing the adoption of big data analytics in the digital transformation era: Case study of Jordanian SMEs. Sustainability, 14(3): 1802. https://doi.org/10.3390/su14031802

[4] Maroufkhani, P., Tseng, M., Iranmanesh, M., Ismail, W.K.W., Khalid, H. (2020). Big data analytics adoption: Determinants and performances among small to medium-sized enterprises. International Journal of Information Management, 54: 102190. https://doi.org/10.1016/j.ijinfomgt.2020.102190

[5] Aldossari, S., Mokhtar, U.A., Ghani, A.T.A. (2023). Factor influencing the adoption of big data analytics: A systematic literature and experts review. SAGE Open, 13(4): 1-25. https://doi.org/10.1177/21582440231217902

[6] Tornatzky, L.G., Fleischer, M. (1990). The Processes of Technological Innovation. Lexington Books.

[7] Raut, R.D., Mangla, S.K., Narwane, V.S., Gardas, B.B., Priyadarshinee, P. (2019). Linking big data analytics and operational sustainability practices for sustainable business management. Journal of Cleaner Production, 224: 10-24. https://doi.org/10.1016/j.jclepro.2019.03.181

[8] Risqi Arifia, A. (2024). Adoption of digitalization in SMEs using the TOE framework and advanced analyses. Journal Economic Business Innovation, 1(2): 180-194. https://doi.org/10.69725/jebi.v1i2.174

[9] Franzen, D.W. (1999). A Bayesian decision model for battle damage assessment. Master's thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, USA. https://scholar.afit.edu/etd/5291/.

[10] McLachlan, S., Dube, K., Hitman, G.A., Fenton, N.E., Kyrimi, E. (2020). Bayesian networks in healthcare: Distribution by medical condition. Artificial Intelligence in Medicine, 107: 101912. https://doi.org/10.1016/j.artmed.2020.101912

[11] Paucar, L.H.G., Bencomo, N., Sutcliffe, A. (2021). Towards technology acceptance: A Bayesian network of soft requirements, the case of the NHS COVID-19 Test and Trace app. In 2021 IEEE International Requirements Engineering Conference Workshops (REW), Notre Dame, IN, USA, pp. 146-152. https://doi.org/10.1109/REW53955.2021.00026

[12] Nedevschi, S., Sandhu, J.S., Pal, J., Fonseca, R., Toyama, K. (2006). Bayesian networks: An exploratory tool for understanding ICT adoption. In Information and Communication Technologies and Development (ICTD), pp. 277-284. https://doi.org/10.1109/ICTD.2006.301865

[13] Tudoran, A.A., Heikkinen, I. (2014). Estimating online reviews adoption: A Bayesian network approach. In Ideas in Marketing: Finding the New and Polishing the Old: Proceedings of the 2013 Academy of Marketing Science (AMS) Annual Conference, Berkeley, CA, USA, pp. 486-495. https://doi.org/10.1007/978-3-319-10951-0_183

[14] Kraus, S., Jones, P., Kailer, N., Weinmann, A., Chaparro-Banegas, N., Roig-Tierno, N. (2021). Digital transformation: An overview of the current state of the art of research. SAGE Open, 11(3): 1-15. https://doi.org/10.1177/21582440211047576

[15] Shahadat, M.M.H., Nekmahmud, M., Ebrahimi, P., Fekete-Farkas, M. (2023). Digital technology adoption in SMEs: What technological, environmental and organizational factors influence in emerging countries? Global Business Review. https://doi.org/10.1177/09721509221137199

[16] Akter, S., Wamba, S.F., Gunasekaran, A., Dubey, R., Childe, S.J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? International Journal of Production Economics, 182: 113-131. https://doi.org/10.1016/j.ijpe.2016.08.018

[17] Awa, H.O., Ojiabo, O.U., Orokor, L.E. (2017). Integrated technology-organization-environment (T-O-E) taxonomies for technology adoption. Journal of

Enterprise Information Management, 30(6): 893-921. https://doi.org/10.1108/JEIM-03-2016-0079

[18] Kitson, N.K., Constantinou, A.C., Guo, Z., Liu, Y., Chobtham, K. (2023). A survey of Bayesian network structure learning. Artificial Intelligence Review, 56(8): 8721-8814. https://doi.org/10.1007/s10462-022-10351-w

[19] Liu, Y., Constantinou, A.C. (2023). Improving the imputation of missing data with Markov Blanket discovery. In The Eleventh International Conference on Learning Representations (ICLR 2023), Kigali, Rwanda. https://openreview.net/forum?id=GrpU6dxFmMN.

[20] Marchena Sekli, G.F., De La Vega, I. (2021). Adoption of big data analytics and its impact on organizational performance in higher education mediated by knowledge management. Journal of Open Innovation: Technology, Market and Complexity, 7(4): 221. https://doi.org/10.3390/joitmc7040221