Proceedings of the 90th Annual International Meeting of the Psychometric Society, Minneapolis, United States (2025).

An Innovative Approach to Latent Underlying Structure by Employing Bayesian Network Analysis

J. Li [†], Z. Lu [†*] and P. Ji [‡]

†Quantitative Methodologies, University of Georgia, Athens, 30602, GA, USA

‡Statistics Department, University of Georgia, Athens, 30602, GA, USA

*Corresponding author. Email: zlu@uga.edu

Abstract

The study examines a Bayesian Network Analysis (BNA) as a competing alternative framework for modeling the latent structure underlying the interrelations between observed, measured variables. A Bayesian Network (BN) is first introduced by outlining its theoretical foundation, structure-learning algorithm, and evaluation criteria. By employing a hill-climbing algorithm with a Gaussian likelihood, a BN structure is learned from the dataset via the Louvain method. Item groupings derived from BNA are evaluated against true classifications using the Adjusted Rand Index (ARI), purity and recovery metrics. The performance of BNA is then examined through both a real-data illustration and a Monte Carlo simulation, with comparisons to traditional Exploratory Factor Analysis (EFA). In the empirical example, BNA and EFA both recovered the general structure of the Big Five personality domains; however, BNA demonstrated slightly higher adjusted agreement alongside lower purity, suggesting that the two approaches may capture different aspects of the latent structure. In the simulation study, where the data were generated from a factor-based model, EFA slightly outperformed BNA under most conditions. However, the performance gap between the two methods became smaller under more complex conditions, and BNA even surpassed EFA in some cases, despite the true latent structure being factor-based. These results suggest that both approaches were able to recover the general structure, although EFA was more effective when the true structure closely followed a latent factor model. Rather than positioning BNA as a replacement for EFA, the findings support its role as a promising complementary approach that may provide additional insights into item-level relationships and relational structure.

Keywords: Classification, clustering, Bayesian Network Analysis (BNA), probabilistic graphical models, exploratory factor analysis (EFA).

1. Introduction

Traditional psychometric techniques, including exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), have been widely used in social and behavioural sciences to understand the latent structure and relationships among observed data. However, prior research also suggests that latent variable models such as EFA rely on assumptions of local independence and common-cause structure, which may be violated in the presence of direct item–item interactions or complex conditional dependencies among items and latent variables (Borsboom, 2008; Epskamp et al., 2018). For example, in personality assessment using instruments such as the Big Five Inventory, items are often assumed to load primarily on a single latent trait (Mammadov et al., 2024). However, in practice, many items exhibit cross-loadings and residual associations that are not fully captured by a simple factor structure. Some items may function as “bridges” between traits, for example, an item reflecting emotional expressiveness may relate to both Extraversion and Agreeableness. These patterns create a web of inter-item relationships that cannot be fully represented by independent latent factors alone. As a result, traditional factor-analytic approaches may oversimplify the structure.

To address this limitation, network-based approaches, including Bayesian network Analysis (BNA) models, provide a flexible framework for representing these dependencies explicitly (Pearl, 1988; Scutari, 2010). BNA constructs graphical models that represent conditional dependencies among variables, enabling more flexiblility in exploring latent trait structures. BNA provides a data-driven approach for revealing underlying item relationships, especially in multidimensional instruments. Bayesian Networks (BNs) have gained prominence due to their ability to model complex conditional dependencies. Moreover, BNs offer advantages such as handling missing data and integrating prior knowledge, which are features valuable in analyzing large-scale inventories. Sun et al. (2014) and Zhou et al. (2014) emphasized BNs' flexibility in diverse contexts, supporting their potential for mapping the nuanced structure of personality instruments.

Despite the increasing interest in BNA within the social and behavioural sciences, the comparison between BNA and traditional factor analysis remains limited. Prior work has demonstrated BNA's effectiveness in uncovering latent constructs and modelling nonlinear relationships (Zhou et al., 2014; Sun et al., 2014), but its utility in aligning observed item structures with theoretical frameworks has not been fully explored.

The purpose of this study is to apply BNA to the domain of factor analysis and evaluate the resulting groupings though Purity, Adjusted Rand Index (ARI), and Recovery. By modeling conditional dependencies, this analysis seeks to advance psychometric methods and explore whether BNA can more accurately recover the true latent dimensions.

2. Methods

2. 1 Bayesian Networks (BNs)

A Bayesian network is represented as a directed acyclic graph (DAG) defined by:

$$G = (V, E)$$

where V is the set of nodes such as items and E is the set of directed edges denoting dependencies. Suppose the data is $D = (X_1, X_2, \dots X_n)$, then using Bayes' theorem, the overall network structure G is identified by maximizing the posterior probability

$$P(G|D) = \frac{P(D|G)\,P(G)}{P(D)} \propto P(D|G)P(G)$$

where $\mathrm{P}(D|G)$ is the likelihood of the data D given the graph structure G, and $P(G)$ is the prior probability of the graph. The joint likelihood over all variables in the data $P(D|G)$ is decomposed as

$$P(D|G) = P(X_1, X_2, \dots X_n|G) = \prod_{i=1}^{n} P(X_i|Pa(X_i))$$

where n is the number of observations, $Pa(X_i)$ is the parent nodes of X_i in G, and $P(X_i|Pa(X_i))$ is the conditional probability of each variable given to its parents. Each variable X_i can be assumed to follow a conditional Normal distribution given to its parents

$$P(X_i|Pa(X_i)) = N(\mu_i, \sigma_i^2)$$

Specially, the conditional probability is

$$P(X_i|Pa(X_i)) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(X_i - \mu_i)^2}{2\sigma_i^2}\right)$$

where $\mu_i = w_i\, Pa(X_i)$ is the conditional mean, w_i is the edge strength, and σ_i^2 is the variance of residuals. This assumption simplifies the representation and computation of dependencies, as the Normal distribution supports analytical solutions for both joint and marginal probabilities.

The existence of an edge between two variables is determined using Bayes' theorem

$$P(E_{ij}|D) = \frac{P(D|E_{ij})\,P(E_{ij})}{P(D)} \propto P(D|E_{ij})P(E_{ij})$$

where $P(D|E_{ij})$ is the likelihood of the data given the edge, and $P(E_{ij})$ is the prior probability of the edge's existence. For edges without explicit priors, the scoring function can be used to rank potential edges, where the score for an edge e_{ij} is defined as

$$Score(E_{ij}) = log\ P(D|E_{ij}) + log\ P(E_{ij})$$

Under the assumption of non-informative $P(E_{ij}) =$ constant, the score can further simplify to

$$Score(E_{ij}) = log\ P(D|E_{ij})$$

2. 2 Algorithm

In the present study, we used a heuristic greedy search algorithm with hill climbing to facilitate estimation of the model, due to its computational efficiency, scalability, and widespread use in the literature (e.g., Heckerman et al, 1995). In addition to this hill-climbing algorithm, other heuristic search strategies, such as simulated annealing, may provide different trade-offs between computational efficiency and the ability to escape local optima. Furthermore, non-heuristic approaches, including maximum likelihood based methods, are also available for network structure learning. However, these approaches are often computationally intensive and may become impractical for moderate to high-dimensional datasets, as the space of possible network structures grows exponentially with the number of variables. For this reason, heuristic search algorithms remain the most commonly used approach in practice.

This algorithm begins with an initial empty graph and iteratively adds or removes edges to maximize the fit between the network structure and the observed data, as determined by the log-likelihood score. After the Bayesian network structure (G*) is learned and its parameters estimated, the edge strengths (w_{ij}) are calculated to quantify the dependency between connected variables X_i and X_j. For continuous variables, the edge strength is defined as the absolute value of the regression coefficient

$$w_{ij} = |\beta_{ij}|$$

where β_{ij} is the coefficient of X_j in the regression model for X_i. Large values of β_{ij} indicate stronger dependencies and suggest that X_i and X_j are likely to belong to the same group. The coefficient is calculated as

$$\beta_{ij} = \frac{Cov(X_i, X_j)}{Var(X_j)}$$

The learned Bayesian network (G*) is then partitioned into groups using a community detection algorithm. The algorithm leverages the computed edge strengths to identify clusters of strongly connected nodes.

The Louvain method is employed to maximize the modularity Q of the graph

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

where A_{ij} is the edge strength w_{ij} between nodes i and j, $k_i = \sum_j A_{ij}$ is the degree of node i, $m = \frac{1}{2}\sum_{i,j} A_{ij}$ is the total edge weight in the graph, and $\delta(c_i, c_j)$ is 1 if nodes i and j belong to the same community, and 0 otherwise. By using w_{ij} as edge weights, the Louvain algorithm optimizes Q to identify groups of nodes that are densely connected based on their estimated dependencies. Higher values of w_{ij} contribute more to the modularity score, indicating stronger connections within the same community. This process ensures that the learned network structure

(as captured by w_{ij} directly informs the grouping of items, providing a data-driven method to partition the items into meaningful communities.

2. 3 Evaluation

The resulting groups obtained were evaluated against the true structure. The metrics used to assess the quality of the learned groups included the Purity, the Adjusted Rand Index (ARI), and Recovery index. The groups were considered well-formed if they exhibited high purity, ARI score, and recovery index.

The purity is the proportion of items correctly grouped. It measures how well predicted clusters align with true class labels.

$$Purity = \frac{1}{N}\sum_{k=1}^{K} \max_{j} |C_k \cap T_j|$$

where N is the total number of items, K is the total number of clusters, C_k are items in cluster *k*, and T_j are items in true class *j*.

ARI is a metric to measure the similarity between the learned grouping and the true structure, accounting for chance. It quantifies the similarity by calculating the following ratio,

$$ARI = \frac{RI - E|RI|}{\max(RI) - E|RI|}$$

where the Rand Index (RI) is computed based on the pair counts,

$$RI = \frac{a + b}{\binom{n}{2}}$$

in which a is the number of pairs of items that are in the same cluster in both clustering; b is the number of pairs of items that are in different clusters in both clustering; *n* is the total number of items; and $\binom{n}{2}$ is the total number of possible pairs of items.

The recovery index summarizes the extent to which the estimated grouping preserved the true pairwise item relationships. It was defined as the proportion of item pairs that the estimated clustering agreed with the true structure. Let z_i denote the true group membership of item i, and $\hat{z}_i$ denote its estimated group membership. The recovery is calculated as

$$\text{Recovery} = \frac{1}{\binom{n}{2}}\sum_{i<j} I[(z_i = z_j \text{ and } \hat{z}_i = \hat{z}_j) \text{ or } (z_i \neq z_j \text{ and } \hat{z}_i \neq \hat{z}_j)],$$

where $I(\cdot)$is an indicator function that equals 1 when the estimated and true pairwise relationships agree and 0 otherwise. Finally, success indicators and BNA-specific diagnostics—the number of detected communities (k), was also recorded so that poor performance could be distinguished from outright estimation failure or systematic under or over-clustering.

3. Real Data Analysis

The data set is the Big Five Inventory-2 (BFI-2), a 60-item measure, corresponding to items P1 through P60, which provides a widely accepted hierarchical framework for capturing five major personality domains and 15 facets (Soto & John, 2017). The sample size is 394 participants.

Due to the relatively small sample size, we treated the data as continuous and normally distributed. We acknowledge that the observed items are ordinal, and BNA is a flexible framework that can accommodate continuous or ordinally-scaled items. We adopted this assumption primarily for computational convenience and consistency with common implementations (e.g., score-based structure with Gaussian likelihood). Treating Likert-type items as continuous is a common approximation in psychometric research, particularly when items have five or

more response categories and distributions are not severely skewed. Under such conditions, the Gaussian assumption provides a reasonable and computationally efficient approximation to the underlying relationships among variables.

We first conducted the BNA to analyze the relationships among the 60 items. To ensure an exploratory analysis, a no-prior knowledge approach was employed for edge probability estimation to allow data to guide the identification of dependencies among items. The resulting classifications were evaluated using ARI and Purity, which quantify the degree of agreement between the BNA -derived groupings and the theoretical Big Five item structure. Second, as a comparison, a traditional EFA was also conducted to benchmark the performance of the BN approach. Different number of factors were tried and finally determined to five, which is in accordance with the theoretical structure of the BFI-2. The resulting factor-based classifications were also evaluated using ARI and Purity.

All analyses were implemented in R (version 4.5.3). Key packages included "psych", "igraph", "bnlearn", "mclust", "aricode", "MASS", "Matrix". Among them, "psych" is for EFA, "igraph" is used in BNA for community detection, "bnlearn" is the core package for BNA framework, "mclust" and "aricode" are to calculate ARI and purity, "MASS" is for the multivariate normal data generation in the simulation study, "Matrix" is to hand the sparse and dense matrices.

3. 1 BNA Results

Using the hill climbing algorithm with the BIC-G score for continuous data, a Bayesian network structure was learned from the dataset, including 60 nodes and 225 directed edges capturing the relationships between items. The average Markov blanket size was found to be 16.33, indicating that on average, each node is conditionally dependent on approximately 16 other nodes in the network. The average neighborhood size (the average number of adjacent nodes) is 7.5, and the average branching factor (the average number of edges per node) is 3.75. This tells us that on average each node is linked to 3.75 other nodes. To quantify the relationships between items, edge strengths w_{ij} were calculated as the absolute values of regression coefficients derived from the fitted Gaussian Bayesian network. Edge strengths ranged from 0.057 (weak) to 0.637 (strong) with a mean value of 0.331, demonstrating a wide variability in the strength of item associations.

The directed graph was transformed into an undirected graph by symmetrizing edge weights. This allowed the application of the Louvain algorithm to identify communities (clusters) of items. The algorithm identified six distinct communities, with a modularity score of Q = 0.35, reflecting a moderate level of cluster separability. The final cluster assignments are listed in Table 1.

Table 1. Classification from BNA

Community 1	P1, P4, P7, P9, P14, P15, P19, P24, P29, P33, P34, P39, P42, P44, P49, P58, P59
Community 2	P6, P11, P16, P21, P31, P46, P56
Community 3	P10, P17, P20, P25, P36, P40, P45, P51, P55
Community 4	P3, P8, P18, P23, P26, P28, P41, P48, P54
Community 5	P5, P13, P30, P32, P35, P38, P43, P50, P53, P60
Community 6	P2, P12, P22, P27, P37, P46, P47, P52, P57

To illustrate the clustering results, a network graph was generated as shown in Figure 1. In this graph, each node corresponds to a BFI-2 item, and the edges between them denote the strength of their conditional relationships. The nodes are color-coded based on their community assignments as determined by the Louvain community detection algorithm, resulting in six distinct clusters. The graph also reveals overlapping relationships between communities, especially in items like those in Community 1 (red), which connect with multiple domains. This suggests that some items may exhibit dual-factor loadings or capture broader personality traits that transcend singular theoretical domains.

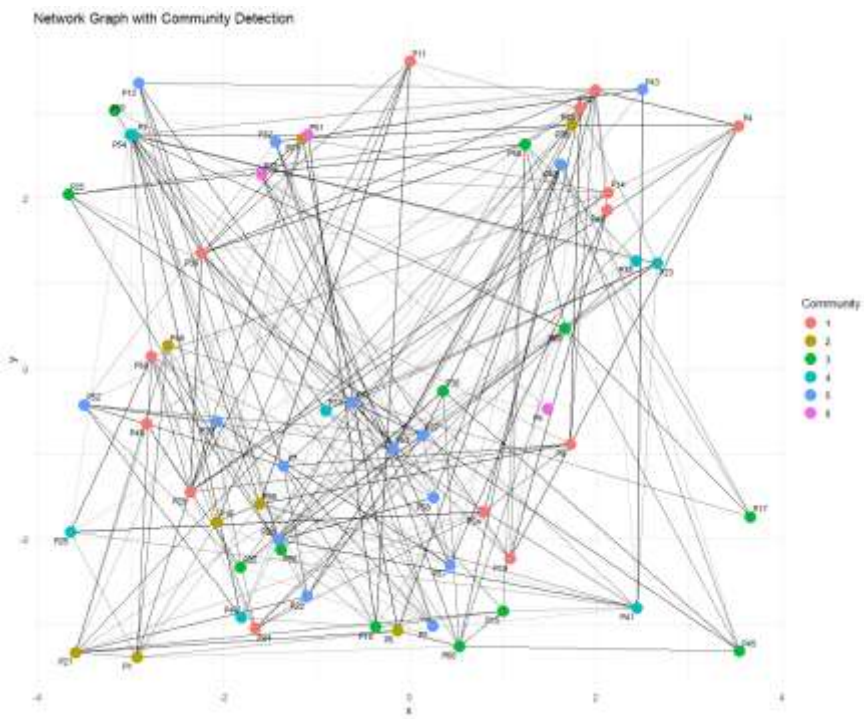

Figure 1. Network graph with community detection

The mapping between the theoretical BFI categories and the predicted clusters suggest some alignment with expectations: all items captured by Communities 2 are from the Extraversion domain, indicate a good recovery of theoretical structure; also, Community 6 has 8 out of 9 items that comes from a same domain Agreeableness; and for Community 3 and 5, there are 11 out of 20 are from domain Open-Mindedness; as for Community 4, 6 out of 9 items are from domain Conscientiousness. However, discrepancies were also noted in terms of the uneven distribution of items across communities. For example, Community 1 captured a relatively diverse set of items across theoretical domains. To evaluate the classification, the purity value was calculated as 0.26 and ARI value was 0.361.

3. 2 EFA Results

To provide a comparative benchmark for the network-based results, an EFA was conducted using the same dataset. Items were assigned to factors based on their highest standardized factor loadings, and the resulting factor composition is reported in Table 2.

Table 2. Classification from EFA

Factor 1	P4, P9, P14, P19, P24, P29, P34, P39, P42, P44, P49, P54, P59
Factor 2	P1, P6, P16, P21, P26, P31, P36, P41, P46, P51, P56
Factor 3	P3, P8, P13, P18, P23, P28, P33, P38, P43, P48, P53, P58
Factor 4	P2, P7, P11, P12, P17, P22, P27, P32, P37, P47, P52, P57
Factor 5	P5, P10, P15, P20, P25, P30, P35, P40, P45, P50, P55, P60

3. 3 Comparison and Conclusions

The factor-based solution served as a conventional reference against which the clustering performance of the Bayesian network could be evaluated. Using the same evaluation criteria applied to the BNA results, the agreement between the EFA-derived factor structure and the theoretical classification was quantified using ARI and

Purity. As shown in Table 3, the EFA solution achieved an ARI of 0.324 and a purity of 0.31. Compared with the BNA results (ARI = 0.361; Purity = 0.26), the factor-analytic solution from EFA was largely comparable, only exhibiting slightly higher purity but slightly lower adjusted agreement. This pattern suggests that while EFA tends to produce clusters with slightly greater internal homogeneity, it may be slightly less effective in capturing the overall relational structure among items.

Table 3. Evaluation Matric for Comparison

Matric	EFA	BNA
ARI	0.324	0.361
Purity	0.31	0.26

The real data analysis results indicate that BNA modeling is capable of recovering theoretically meaningful structure. Rather than positioning BNA as a replacement for traditional models, the results support its role as a supplementary analytic framework that can enhance construct interpretation and diagnostic resolution. By identifying distinct communities through the Louvain algorithm, BNA demonstrated slightly higher adjusted agreement, albeit with lower purity. These findings suggest that BNA may be effective in capturing the overall relational structure among items, especially in contexts characterized by complex item interdependencies or less clearly delineated latent structures.

4. Monte Carlo Simulation

To further evaluate the performance of BNA and EFA recovering a known latent structure, a Monte Carlo simulation study was conducted for a more systematic comparison between both methods under various conditions.

4. 1 Data Generation

In the simulation, we assumed the true underlying structure had five dimensions, consistent with the Big Five framework of the empirical illustration. To keep the simulation feasible yet with meaningful and substantively interpretable manipulated factors, the number of items per factor was fixed at 3 and the sample size was fixed at $N = 200$ across conditions. Data were generated for $m = 5$ latent dimensions with 3 items per dimension, resulting in $p = 15$ observed variables. The population covariance $\mathbf{\Sigma}$ is a 15×15 matrix, which can be calculated as $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}$, where $\mathbf{\Lambda}$ is the 15×5 factor loading matrix, $\mathbf{\Phi}$ is the 5×5 latent factor covariance matrix, and $\mathbf{\Theta}$ is the 15×15 residual covariance matrix. The matrix was adjusted to the nearest positive-definite form to ensure that sampling could proceed under all simulated conditions. Continuous multivariate normal data were first drawn from this covariance structure, and then transformed into five-category Likert-type responses using quintile-based cuts. This produced one observed ordinal dataset per replication. The same observed dataset was used for both EFA and BNA. For comparison, EFA was specified to extract five factors (K = 5) and BNA was constrained to partition to five communities (K = 5). Again, all analyses were implemented in R (version 4.5.3). In addition to the key packages used in real data analysis, "MASS" is for the multivariate normal data generation and "Matrix" is to hand the sparse and dense matrices.

The four manipulated condition factors were cross-loading magnitude, factor correlation, residual-correlation strength, and residual-correlation density. For $\mathbf{\Lambda}$, each item has a primary loading on its home factor and a cross-loading on a non-home factor. By using cross-loadings, it violated the simple structure design (Markos & Tsigilis, 2024). The primary loading was fixed at .45 for all items. Cross-loadings were varied from a relatively mild level at .10 to a more salient level at .30. In practice, a cross-loading of .10 represents limited secondary association, whereas .30 represents a clearly more interpretable secondary loading and therefore a stronger challenge to clean factor separation. For $\mathbf{\Phi}$, the factor correlations varied from a moderate value at .30 to a relatively high correlation at .60 (Golino et al., 2020). The correlation of .60 provided a more difficult condition in which latent dimensions were still distinguishable but much less separable. For $\mathbf{\Theta}$, the residual matrix was manipulated by varying the residual-correlation (.25 and .45) and the residual-correlation density/ proportion (.167 and .333) to increase the item interdependence complexity suggested by Christensen et al. (2023). Residual correlation indicates how strongly two items were associated after the latent factor structure was imposed, and residual density indicated the proportion of the item-level dependencies across the test. In total, there are $2 \times 2 \times 2 \times 2 = 16$ conditions as shown in Table 4.

Conditions became more complex with stronger cross-loadings, higher factor correlations, and greater residual dependence, making recovery of the true structure more challenging. For each condition, 100 replications were generated, yielding 1,600 simulated datasets. To illustrate the population structure used in the simulation, the factor model under Condition 1 is shown in Figure 2.

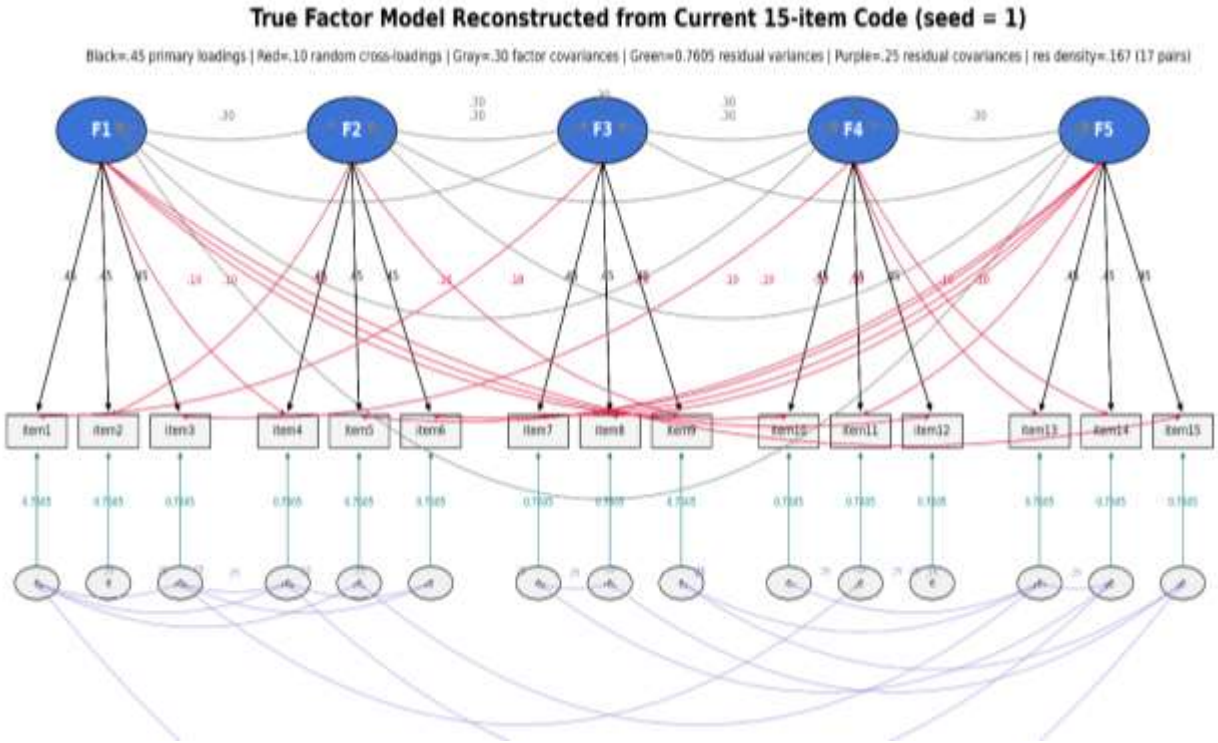

Figure 2. Factor Model Under Condition 1

Table 4. Simulation Design

Condition	Primary loading	Cross-loading	Factor correlation	Residual correlation	Residual density
C1	0.45	0.100	0.300	0.250	0.167
C2					0.333
C3				0.450	0.167
C4					0.333
C5			0.600	0.250	0.167
C6					0.333
C7				0.450	0.167
C8					0.333
C9		0.300	0.300	0.250	0.167
C10					0.333
C11				0.450	0.167
C12					0.333
C13			0.600	0.250	0.167
C14					0.333
C15				0.450	0.167
C16					0.333

4. 2 Results and Conclusions

Table 5 summarizes the results across 16 simulation conditions. Overall, EFA and BNA performed almost the same. The differences between them are very small, with EFA slightly better than BNA. For both methods, performance tended to decline as the conditions became more complex. This pattern was especially clear for ARI, where EFA exceeded or matched BNA across conditions, and for Recovery, where EFA was usually slightly higher than BNA. A more nuanced pattern emerged for Purity. Although EFA usually showed higher purity than BNA, its advantage was not constant across conditions. Under relatively simple settings, the purity gap was more noticeable. For example, in C1 and C2, EFA exceeded BNA by .036 and .048, respectively (.613 vs. .577; .617 vs. .569). However, as the conditions became more complex, this gap narrowed substantially. In C7 and C8, the purity difference shrank to .000 and .004, respectively (.519 vs. .519; .526 vs. .522). In C16 under higher complexity condition, BNA slightly surpassed EFA (.525 vs. .521).

Table 5. Simulation Results

Condition	Purity		ARI		Recovery	
	EFA	BNA	EFA	BNA	EFA	BNA
C1	0.613	0.577	0.215	0.149	0.786	0.770
C2	0.617	0.569	0.211	0.137	0.783	0.763
C3	0.554	0.525	0.111	0.061	0.766	0.734
C4	0.569	0.545	0.125	0.091	0.767	0.746
C5	0.559	0.537	0.119	0.081	0.758	0.753
C6	0.561	0.541	0.111	0.084	0.754	0.748
C7	0.519	0.519	0.063	0.045	0.753	0.730
C8	0.526	0.522	0.065	0.050	0.750	0.738
C9	0.585	0.548	0.162	0.098	0.771	0.753
C10	0.593	0.552	0.173	0.094	0.771	0.751
C11	0.541	0.518	0.088	0.040	0.758	0.731
C12	0.542	0.530	0.096	0.068	0.759	0.743
C13	0.531	0.514	0.078	0.041	0.747	0.738
C14	0.542	0.534	0.092	0.072	0.750	0.747
C15	0.519	0.493	0.053	0.042	0.747	0.722
C16	0.521	0.525	0.059	0.059	0.741	0.744

One reason to explain the slight advantage for EFA over BNA across conditions is the data-generating mechanism was a factor model rather than network based. The population covariance structure was generated from a latent factor model. As a result, the simulated data naturally favored EFA. By contrast, BNA was required to infer conditional dependency patterns from the same observed data and then recover communities from the learned graph, making the recovery task more indirect and therefore more difficult.

5. Discussion, Limitations and Future Directions

This study examined BNA as a competing alternative framework for recovering a latent structure and compared its performance with EFA using both a real-data illustration and a Monte Carlo simulation. In the empirical example, BNA showed slightly higher adjusted agreement but lower purity than EFA, suggesting that Bayesian network modeling is capable of recovering latent structure, although the two approaches may emphasize different aspects of structure recovery. In the simulation study, EFA slightly outperformed BNA, indicating that when the true latent structure follows a factor-based model, EFA is more effective in reproducing that structure. However, the simulation results also suggested that the performance gap between the two methods became smaller under several more complex conditions, and BNA even surpassed EFA (under C 16), despite the true latent structure being factor-based.

Several limitations of the present simulation point to future directions of research. First, the data-generating process assumed a factor-based structure, which may inherently favor EFA. Second, both EFA and BNA were

constrained to recover five groups, which provided comparability between the two methods but also limited the extent to which the fully exploratory behavior of each method could be evaluated. Third, the relatively small number of items (three) per factor and the modest sample size (200) may have limited BNA's ability to form clearly separable communities. Fourth, the residual dependence in the simulation may introduced more noise than as meaningful network structure, potentially disadvantaging BNA. In addition, the multistep nature of the BNA procedure may have more cumulative errors relative to the more direct estimation in EFA.

Future research should address these limitations by expanding the simulation design. For example, studies should consider data generated from network-based structures to provide a more appropriate context for evaluating BNA. Increasing the sample size and the number of items per factor may improve BNA's performance. Further work may also examine the role of residual dependence or the item-level conditional relationships in network structure. In addition, exploring alternative algorithms may be valuable, as different search strategies could potentially influence the recovered network structure. Extensions to discrete or ordinal BNs are also available, for example, through multinomial likelihoods or conditional probability tables, which allow the modeling of categorical or ordinal data directly. Finally, sensitivity analyses across varying item characteristics (Li & Lu, 2026) will help to assess the robustness of both approaches. Together, these extensions would provide a more comprehensive understanding of when BNA can serve as a complementary method to traditional factor analysis.

Acknowledgement

We sincerely thank the Chief Editor, Dr. Jee-Seon Kim, for the opportunity to submit our manuscript. We also gratefully acknowledge the Co-Editor, Dr. Jeffrey Harring, for his thoughtful and constructive comments, as well as for his time and support.

Funding Statement

None

Competing Interests

None

References

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of clinical psychology*, 64(9), 1089-1108.

Christensen, A. P., Garrido, L. E., & Golino, H. (2023). Unique variable analysis: A network psychometrics method to detect local dependence. *Multivariate Behavioral Research*, 58(6), 1165-1182.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior research methods*, 50(1), 195-212.

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological methods*, 25(3), 292.

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3), 197-243.

Li, J., & Lu, Z. (2026). Evaluating Psychometric Clustering Methods: A Machine-Learning Comparison of EFA and NCD. *Computation*, 14(4), 80.

Mammadov, S., Wang, S., & Lu, Z. (2024). Personality types and their associations with psychological resilience, coping with stress, and life satisfaction among undergraduate students: A latent profile analysis approach. *Personality and Individual Differences*, 222, 112599.

Markos, A., & Tsigilis, N. (2024). Dimensionality assessment in ordinal data: a comparison between parallel analysis and exploratory graph analysis. *Frontiers in Psychology*, 15, 1359111.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Scutari, M. (2010). Learning Bayesian networks with the bnlearn R package. *Journal of statistical software*, 35, 1-22.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69-81.

Sun, Y., He, M., & Yang, L. (2014). Comprehensive influence model of preschool children's personality development based on the Bayesian Network. In *Abstract and Applied Analysis* (Vol. 2014, No. 1, p. 267953). Hindawi Publishing Corporation.

Zhou, Y., Fenton, N., & Neil, M. (2014). Bayesian network approach to multinomial parameter learning using data and expert judgments. *International Journal of Approximate Reasoning*, 55(5), 1252-1268.