# ORIGINAL RESEARCH ARTICLE

# Bayesian networks for inferring the relationship between individual behavior and social influence: A case study of early 20th century British travels in China

**Meijie Ding**

*Faculty of Education, Language, and Communication, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, 94300, Malaysia*

**\* Corresponding author:** Meijie Ding, meijiebfnf@outlook.com

## ABSTRACT

This study investigated the possibility of applying the Bayesian networks (BNs) in analyzing the relationship between individual behavior and social influence among early 20th-century British travelers in China. While historical studies have provided valuable details about social interactions, existing research using such studies has shown limitations in quantifying and analyzing complex relationships. This study attempts to address this gap by employing Bayesian networks (BNs) to construct a framework for modeling the probabilistic relationships between various factors influencing the travel patterns of British travelers in China in the early 20th century. These factors include political climate, economic considerations, and cultural interactions, which are sourced through historical studies, travel diaries, and other contemporary sources. The performance of the proposed Bayesian network model is evaluated using established statistical methods, including confusion matrices, cross-validation, and sensitivity analysis (SA). The results have shown the significance of the chosen model in analyzing the complex relationship selected analysis.

*Keywords:* individual behavior and social influence; social science; Bayesian networks; cross-validation; sensitivity analysis

## 1. Introduction

The early 20th century witnessed a significant period in global history characterized by intensified international trade exchanges and colonial expansion[1], and this period generated a wealth of data that had emerged from the travelogues of British explorers in China. These data contain documentation of various interactions, including economic exchanges and complex cultural and social engagements[2]. Such interactions are analyzed to obtain the perspectives that provide valuable insight into the Historical Context (HC); such perspectives still influence contemporary global dynamics. However, existing models used for the analysis use descriptive models to gather insights, which cannot include subtle relations among different social factors. This limitation must be addressed by developing a quantitative model that is capable of understanding the complex multidimensional influence on an individual's behavior with respect to the HC[3].

Current models that are used for the examination of HC are more dependent on linear frameworks, which are not capable of handling the complexities of social networks[4]. The Traditional statistical methods have

shown limited ability to model the non-linear and probabilistic nature of social interactions and their impact on individual behavior[5]. These limitations are attributed to the historical and social dynamics involving multifaceted interactions influenced by various factors. Additionally, the existing models cannot represent the likelihood of social influence and individual choice preference in response to evolving social, political, and cultural environments[6]. To address these limitations, this work has chosen Bayesian Networks (BN), a model that employs interdependent variables to handle uncertain and incomplete data, which makes it a good choice for historical social studies. By using BN, the work further explores how individual behaviors were shaped based on the social dynamics of the historical time[7].

Using the BN model, the work explores the complex interactions between British travelers and the socio-cultural environment of China in the early 20th century. The work starts with data collection through various sources, including travel diaries, letters, official records, and contemporary accounts[8]. Using the data, the probabilistic relationships among the identified factors are built using the BN model. The conditional probability scores for each factor are measured, highlighting the crucial factors influencing travel behavior during that time. The constructed model is evaluated using statistical techniques like confusion matrices, cross-validation, and Sensitivity Analysis (SA). The results from this analysis prove the BN model's applicability in historical social studies.

The paper is structured as follows: Section 2 presents the background, Section 3 presents the model, Section 4 analyzes the results, and Section 5 concludes the work.

# 2. Background

The early 20th century is identified as a period of significant international interaction and explorations of travelers to many foreign countries. This was when many British travelers had shown interest and traveled to the eastern part of the globe, particularly China. The increase in interest among British nationals is primarily attributed to the trade and the sharing of ideas, beliefs, and social norms[9-10].

A better understanding of individual behaviors that had raised interest among British travelers during that time helped to understand better the HC associated with it. To assist this objective, a method based on a probabilistic graphical model known as Bayesian Networks is chosen to understand the relationships between individual behavior and social influence[11-13].

## 2.1. Bayesian Networks (BN)

BN are graphical models that represent a set of variables and their conditional dependencies via a directed acyclic graph (DAG). These networks are used to model the probabilistic relationships among variables. Each node in the graph represents a variable, and the edges relate to the probabilistic dependencies between these variables[14-16]. The strength of these dependencies is quantified using conditional probabilities. BN's strength is its ability to combine prior knowledge (in the form of probability distributions) with new data to update beliefs about the state of the world using Bayes's Theorem. In the current study, the applicability of BN is particularly useful for modeling complex systems with many interdependent variables[18-21].

## 2.2. Historical context

The early 20th century was a period of significant transformation for both Britain and China that was marked by rapid industrialization, political upheaval, and evolving international relations. In Britain, that time was characterized by the tail end of the Victorian era, which transitioned into the Edwardian period right after the aftermath of the First World War. This period saw Britain at the peak of its imperial power, which had an extensive empire that was often referred to as an empire where the sun never set down. Furthermore, British

society was undergoing considerable change related to technological advancements like the steam engine and telegraph, and the social structures were being challenged by emerging political movements, including suffragism and labor rights. At the same time, the Qing Dynasty, which had been ruling China for centuries, was in its final stages in China, and in 1911, the Republic of China was formed. During this period, China was engulfed with internal strife, efforts to modernize, and increased foreign influence and intervention through unequal treaties and establishing foreign concessions.

British travels to China during this time were motivated by different factors, including trade, diplomacy, missionary work, and tourism. Remarkably, the trade related to tea, silk, and porcelain has primarily influenced the traveler's interest in China. Right after the Opium Wars (1839-42 and 1856-60), many treaty ports were created, increasing travel. Cities such as Shanghai and Hong Kong had seen a steady increase in the influx of expatriates from Britain. Further, the invention of steamships and the opening of the Suez Canal in 1869 reduced the travel time between Europe and Asia; this, too, contributed to the increase in the frequency of British travels to China. This accessibility boosted trade and enabled various persons, including scholars, artists, and tourists, to visit China.

Such British travelers often documented their experiences in China through writings, photographs, and collections of art and artifacts, which contributed to a fascination with Chinese culture in Britain. However, these interactions were defined by a complex relationship of imperialism, cultural exchange, and mutual curiosity that reflected the broader dynamics of East-West relations during this period.

# 3. Methodology

## 3.1. Data sources

This study uses different HC data sources to analyze the BN model. The data sources cover the entire spectrum of the socio-cultural dynamics of early 20th-century British travels in China. The primary categories of data sources utilized include:

**(1) Historical records:** Official documents, governmental reports, and archives provide information about the political and social context of the era. These records contain details about diplomatic relations, trade policies, and other factors influencing travel patterns.

**(2) Travel diaries:** Personal diaries of British travelers provide firsthand accounts of detailed descriptions of daily activities, experiences, perceptions, and interactions with local Chinese communities.

**(3) Letters:** Correspondences between travelers and their families, friends, or associates offer an understanding of the personal feelings attached to the experiences and thoughts of travelers about events or situations encountered in China.

**(4) Contemporary accounts:** Newspaper articles, books, and journals written during the period provide context and commentary on British travels in China. These are written by journalists, scholars, or other travelers, and such writings offer external perspectives on the behaviors and social influence of British travelers, as well as general public perception.

**(5) Photographs and artifacts:** Photographs and artifacts from that era offer additional dimensions that reflect the interactions, activities, and environments experienced by British travelers. Artifacts like souvenirs travelers collect can also indicate cultural exchanges and influences.

These varied sources enable a comprehensive and nuanced construction of the BN, capturing both the macro-level influences of social and political forces and the micro-level dynamics of individual decision-making and behavior.
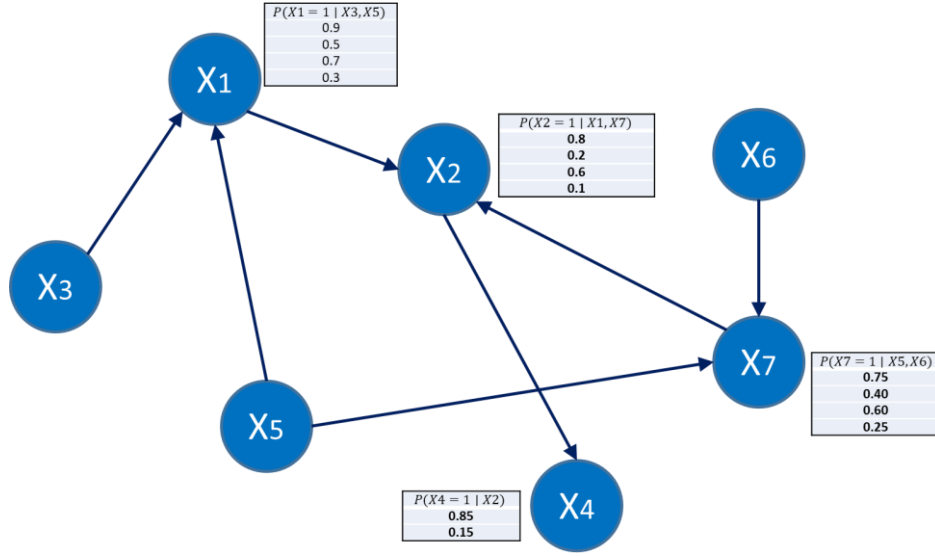
**Figure 1.** BN for the case study.

## 3.2. Bayesian network model

In this section, we present a BN model for the chosen case study (**Figure 1**). This model is structured to illustrate the probabilistic relationships between various factors influencing travel behaviors, using standard notations found in research articles. Let's represent the critical variables as nodes in the BN. For instance,

- $X_1$ for 'Motivation for Travel'

- $X_2$ for 'Duration and Frequency of Travel'

- $X_3$ for 'Social Background of Travelers'

- $X_4$ for 'Interaction with Chinese Culture'

- $X_5$ for 'Political Climate in Britain'

- $X_6$ for 'Political Climate in China'

- $X_7$ for 'Perceived Safety and Risks'

The edges represent dependencies between variables. For example, an edge from $X_5$ ('Political Climate in Britain') to $X_1$ ('Motivation for Travel') indicates a dependency, which can be represented as $P(X_1 \mid X_4)$, the probability of $X_1$ given $X_4$. The entire network is represented by a joint probability distribution, decomposed into a product of conditional probabilities, EQU (1).

$$P(X_1, X_2, X_3, \dots, X_n) = \prod_{i=1}^{n} P\big(X_i \mid \text{Parents}(X_i)\big) \tag{1}$$

where Parents $(X_i)$ represents the set of nodes having a direct influence on $X_i$. For each node, a conditional probability table (CPT) is formulated in EQU (2). For example, for the node 'Motivation for Travel' $(X_1)$, the CPT might look like:

$$- P(X) = \begin{cases} p_{11} & \text{If Political Climate is Stable} \\ p_{12} & \text{If Political Climate is Unstable} \end{cases} \tag{2}$$

where $p_{11}$ and $p_{12}$ are the probabilities of a particular motivation given the state of the political climate. To capture dynamic changes over time, we can introduce time-indexed variables, such as $X_{1,t}$ for 'Motivation for Travel' at time $t$. The temporal dependencies can be represented as $P\big(X_{1,t+1} \mid X_{1,t}, X_{2,t}, \dots\big)$.

4

**Table 1.** The conditional probability table (CPT) for the node 1.

| X3: Social Background of Travelers | X5: Political Climate in Britain | $P(X1 = 1 \mid X3, X5)$ |
|:---:|:---:|:---:|
| 1 | 1 | 0.9 |
| 1 | 0 | 0.5 |
| 0 | 1 | 0.7 |
| 0 | 0 | 0.3 |

**Table 1** reads as follows: for instance, if a traveler comes from a particular social background ( $X3 = 1$ ) and is in a specific political climate in Britain ( $X5 = 1$ ), then there is a high probability (0.9) that their motivation for travel ($X1$) is substantial.

**Table 2.** The conditional probability table (CPT) for the node 2.

| X1: Motivation for Travel | X7: Perceived Safety and Risks | $P(X2 = 1 \mid X1, X7)$ |
|:---:|:---:|:---:|
| 1 | 1 | 0.8 |
| 1 | 0 | 0.2 |
| 0 | 1 | 0.6 |
| 0 | 0 | 0.1 |

**Table 2** indicates that, for example, if a traveler has a high motivation for travel ( $X1 = 1$ ) and perceives safety and risks favorably ( $X7 = 1$ ), then there is a high probability (0.8) that the duration and frequency of their travel ( $X2$ ) will be more significant. Conversely, if motivation and perceived safety are low ( $1 = 0$ and $X7 = 0$ ), the probability that travel duration and frequency are high is much lower (0.1).

**Table 3.** The conditional probability table (CPT) for the node 4.

| X2: Duration and Frequency of Travel | $P(X4 = 1 \mid X2)$ |
|:---:|:---:|
| 1 | 0.85 |
| 0 | 0.15 |

**Table 3** suggests that travelers with more extended and more frequent travels ( $X2 = 1$ ) have a higher probability (0.85) of having significant interaction with Chinese culture ( $X4 = 1$ ). In contrast, those with shorter and less frequent travels ($X2 = 0$ ) have a much lower probability (0.15) of such interaction.

**Table 4.** The conditional probability table (CPT) for the node 7.

| X5: Political Climate in Britain | X6: Political Climate in China | $P(X7 = 1 \mid X5, X6)$ |
|:---:|:---:|:---:|
| 1 | 1 | 0.75 |
| 1 | 0 | 0.40 |
| 0 | 1 | 0.60 |
| 0 | 0 | 0.25 |

**Table 4** indicates, for instance, that if both the political climate in Britain ( $X5$ ) and in China ( $X6$ ) are favorable, then there is a high probability (0.75) of perceiving high safety and low risks ($X7$). Conversely, if both political climates are unfavorable, the likelihood of perceiving high safety and low risks significantly decreases to 0.25.

By altering the values of particular nodes, we can perform predictive analysis. For instance, modifying $X_4$ to reflect a change in the political climate and observe the resultant probabilities in $X_1, X_2, \dots, X_n$ . The SA

is approached by observing the changes in the output probabilities with marginal changes in the input probabilities, quantified as: $\frac{\partial P(X_i)}{\partial P(X_j)}$ indicating how sensitive $X_i$ is to changes in $X_j$. Further, the Probabilistic Interpretation allows for interpreting how varying degrees of social influences impact travel behavior. For instance, calculating the conditional probabilities such as $P(X_2 \mid X_3, X_4)$ offers insights into how the social background and political climate influence travel duration and frequency.

# 4. Analysis

For assessing the performance and accuracy of a BN model in the context of a case study, the following statistical measures were measured:

## 4.1. Confusion matrix and accuracy metrics

The confusion matrix in Table 5 assesses the model's classification in terms of TP, FP, FN, and TN. In this analysis, the BN model correctly predicted 120 instances as positive (TP) and 305 cases as negative (TN). However, it labeled 30 instances as positive when they were negative (FP) and 45 as negative when they were positive (FN).

To translate the confusion matrix into a more comprehensive evaluation, three key metrics, accuracy, precision, and recall, were used to evaluate the model. The accuracy is the sum of TP and TN among all predictions for which the BN scored 85%. Precision measures the proportion of TP among optimistic predictions for which the BN had scored 80%. For Recall, which is the proportion of TP that were correctly identified, the BN had scored 72.7%.

**Table 5.** Confusion matrix.

| Predicted/Actual | Positive (1) | Negative (0) |
|---|---|---|
| Positive (1) | 120 | 30 |
| Negative (0) | 45 | 305 |

**Accuracy metrics:**

- Accuracy: $\frac{120+305}{120+30+45+305} = 0.85(85\%)$

- Precision: $\frac{120}{120+30} = 0.8(80\%)$

- Recall: $\frac{120}{120+45} = 0.727$ (Approximately 72.7% )

The analysis shows that the accuracy of the BN model is comparatively better than that of precision and recall. This indicates that the model has considerable space for enhancement to reduce the false negatives to increase recall.

## 4.2. Cross-validation

The results of the 5-fold cross-validation process are shown in **Table 6** and **Figure 2**. In this cross-validation analysis, the input data is divided into five distinct folds, and the model is trained and tested five times, with one different data fold used for testing and the remaining folds for training.

**Table 6.** Fold vs. accuracy.

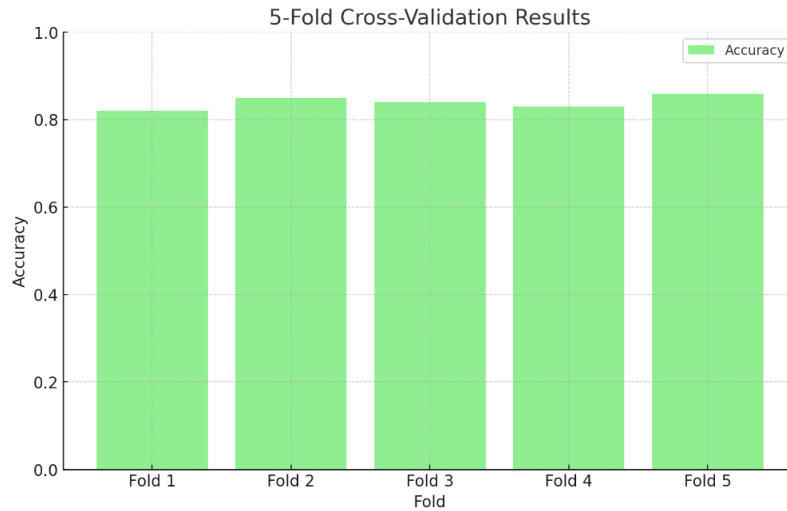| Fold | Accuracy |
|---|---|
| **Fold 1** | 0.82 |
| **Fold 2** | 0.85 |
| **Fold 3** | 0.84 |
| **Fold 4** | 0.83 |
| **Fold 5** | 0.86 |
| **Mean** | 0.85 |
| **Standard Deviation** | 0.0141 |



**Figure 2.** Cross-validation results.

The mean accuracy across all folds was calculated to be 0.84, which shows the model's consistent and reliable performance. The model also showed a low standard deviation of 0.0141 in accuracies across folds, proving the model has minimal variation and is robust in prediction across different data folds.

### 4.3. ROC curve analysis

The ROC curve is a graphical representation of a classification model's diagnostic ability as its discrimination threshold is varied. The curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings.
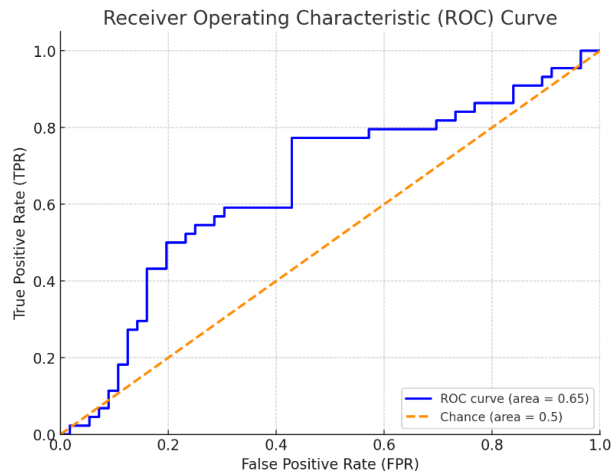


**Figure 3.** ROC curve.

7

In the generated ROC curve in **Figure 3**, the model has an AUC of 0.65. The AUC value ranges from 0 to 1, where an AUC of 1 indicates a perfect model that makes all predictions correctly. An AUC value of 0.5 suggests performance no better than random chance, and an AUC less than 0.5 implies a model performing worse than random chance, which might indicate that the model is inversely predicting the outcomes. The AUC of 0.65 for this model suggests that it can reasonably distinguish between the two classes (e.g., high versus low interaction with Chinese culture). However, there is significant room for improvement, as the model is not far from the AUC of 0.5, representing a random guess.

Given the AUC score, the model is considered to have moderate predictive power. It can correctly rank a randomly chosen positive instance higher than a randomly chosen negative instance with a probability of 0.65. For our case study, this means that while the model has some predictive capabilities, it may not be highly reliable in all scenarios. It could be more effective in some areas of prediction, such as identifying travelers with high motivation for cultural interaction, but less effective in others.

## 4.4. Sensitivity analysis

The analysis explores the extent to which changes in input variables influence the likelihood of increased travel frequency, demonstrating the model's responsiveness to different socioeconomic and technological factors.

**Table 7.** Sensitivity analysis.

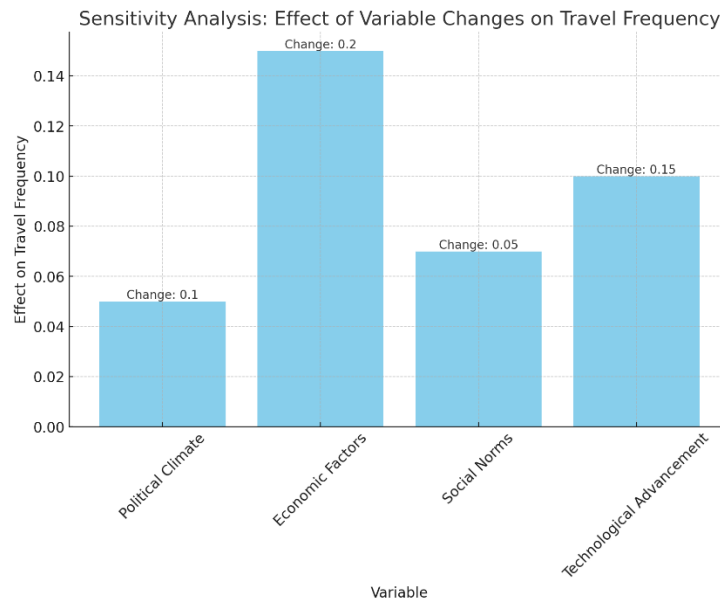| Variable | Change | Effect on Travel Frequency |
|---|---|---|
| **Political Climate** | 0.10 | 0.05 |
| **Economic Factors** | 0.20 | 0.15 |
| **Social Norms** | 0.05 | 0.07 |
| **Technological Advancement** | 0.15 | 0.10 |



**Figure 4.** Sensitivity analysis.

In **Table 7** and **Figure 4**, for the political climate, a change of 0.10 (presumably indicating a 10% improvement in political stability or relations) leads to a 0.05 (or 5%) increase in travel frequency. This suggests a moderate sensitivity of travel frequency to political conditions, reflecting how diplomatic and political environments could impact travel decisions. For economic factors, a more substantial change of 0.20

(indicative of a 20% improvement in economic conditions or opportunities) results in a significant 0.15 (15%) increase in travel frequency. This indicates a strong correlation between economic conditions and travel behavior, highlighting the importance of financial incentives in influencing travel. For social norms, a minor adjustment in social norms by 0.05 leads to a 0.07 increase in travel frequency. This 7% increase, resulting from a relatively small change, underlines the sensitivity of travel behavior to social factors, possibly reflecting changes in societal attitudes towards travel or interactions with different cultures. For technological advancement, a 0.15 (15%) advancement in technology correlates with a 0.10 (10%) increase in travel frequency, showing that improvements in transportation or communication would encourage travel.

# 5. Conclusion

The application of Bayesian Networks (BN) in this study has provided valuable insights into the complex relationship between individual behavior and social influence during the early 20th-century British travels in China. By employing different historical data for analysis using a probabilistic graphical model, the research has successfully quantified and analyzed the dynamics of this historical period. The findings revealed how various factors, such as political climate, economic conditions, and technological advancements, influenced British travel behaviors. The models were experimented with using different statistical evaluations, and the results have shown the potential of BN in analyzing the historical and social science aspects. This analysis using BN provides an understanding of the past and also provides ways for applying advanced statistical methods in the exploration of historical and cultural phenomena. The study has shown the significance of interdisciplinary approaches in understanding the complexity of human behavior and social interactions across different historical epochs.

# Conflict of interest

The authors declare no conflict of interest.

# References

1.  Veracini, L. (2022). Colonialism: a global history. Routledge.
2.  Siu, H. F. (2016). Tracing China: A forty-year ethnographic journey. Hong Kong University Press.
3.  Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., ... & Ungar, L. H. (2021). Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. Psychological Methods, 26(4), 398.
4.  Quintana, R. (2023). Embracing complexity in social science research. Quality & Quantity, 57(1), 15-38.
5.  Mehmetoglu, M., & Jakobsen, T. G. (2022). Applied statistics using Stata: a guide for the social sciences. Sage.
6.  Rounsevell, M. D., Arneth, A., Brown, C., Cheung, W. W., Gimenez, O., Holman, I., ... & Shin, Y. J. (2021). Identifying uncertainties in scenarios and models of socio-ecological systems in support of decision-making. One Earth, 4(7), 967-985.
7.  J. Zheng, S. He, and C. Lv, Application of Bayesian Network in Effectiveness Evaluation of High-Level Construction of Specialized Group, 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC ), Beijing, China, 2022, pp. 597-601, doi: 10.1109/IAEAC54830.2022.9929943.
8.  S. T. Low, M. S. Mohamad, S. Omatu, L. E. Chai, S. Deris and M. Yoshioka, Inferring gene regulatory networks from perturbed gene expression data using a dynamic Bayesian network with a Markov Chain Monte Carlo algorithm, 2014 IEEE International Conference on Granular Computing (GrC), Noboribetsu, Japan, 2014, pp. 179-184, doi: 10.1109/GRC.2014.6982831.
9.  Y. Miyakoshi and S. Kato, Facial emotion detection considering partial occlusion of face using Bayesian network, 2011 IEEE Symposium on Computers & Informatics, Kuala Lumpur, Malaysia, 2011, pp. 96-101, doi: 10.1109/ISCI.2011.5958891.
10. M. Nouh and J. R. C. Nurse, Identifying Key Players in Online Activist Groups on the Facebook Social Network, 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 2015, pp. 969-978, doi: 10.1109/ICDMW.2015.88.

11. A. Alsayat and H. El-Sayed, Social media analysis using optimized K-Means clustering, 2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA), Towson, MD, USA, 2016, pp. 61-66, doi: 10.1109/SERA.2016.7516129.
12. E. Otte and R. Rousseau, Social network analysis: a powerful strategy also for the information sciences, Journal of Information Science, vol. 28, no. 6, pp. 441-453, 2002.
13. T. C. Haas and S. M. Ferreira, Federated databases and actionable intelligence: using social network analysis to disrupt transnational wildlife trafficking criminal networks, Security Informatics, vol. 4, no. 2, 2015.
14. T. Opsahl, F. Agneessens and J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths, Social Networks, vol. 32, no. 3, pp. 245-251, 2010.
15. David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al., Life in the network: the coming age of computational social science in Science, New York, NY, vol. 323, no. 5915, pp. 721, 2009.
16. Kevin Lerman, Ari Gilder, Mark Dredze and Fernando Pereira, Reading the markets: Forecasting public opinion of political candidates by news analysis, Proceedings of the 22nd International Conference on Computational Linguistics-Volume, vol. 1, pp. 473-480, 2008.
17. Y. Fan and Q. Peng, Inferring gene regulatory networks based on spline regression and Bayesian group lasso, 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Shanghai, China, 2016, pp. 39-42, doi: 10.1109/SNPD.2016.7515875.
18. C. Wang, X. Gao and X. Li, An Interpretable Deep Bayesian Model for Facial Micro-Expression Recognition, 2023 8th International Conference on Control and Robotics Engineering (ICCRE), Niigata, Japan, 2023, pp. 91-94, doi: 10.1109/ICCRE57112.2023.10155596.
19. J. Zheng, Z. Wei and W. Zheng, The casual inference of road traffic accidents based on the Bayesian network optimization, 2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI), Prague, Czech Republic, 2023, pp. 185-193, doi: 10.1109/ACEDPI58926.2023.00044.
20. Q. Wang et al., Fault Intelligent diagnosis of reversible pumped storage Unit based on Bayesian networks and counterfactual reasoning, 2023 8th Asia Conference on Power and Electrical Engineering (ACPEE), Tianjin, China, 2023, pp. 2018-2022, doi: 10.1109/ACPEE56931.2023.10135793.
21. N. Sharmin, S. Roy, A. Laszka, J. Acosta and C. Kiekintveld, Bayesian Models for Node-Based Inference Techniques, 2023 IEEE International Systems Conference (SysCon), Vancouver, BC, Canada, 2023, pp. 1-8, doi: 10.1109/SysCon53073.2023.10131168.