

RESEARCH ARTICLE

Data Science and Analytics

Network analysis of tuberculosis spread with forecasting and technique for order preference by similarity to ideal solution

PNM Perera and KKKR Perera*

Department of Mathematics, Faculty of Science, University of Kelaniya, Kelaniya.

Submitted: 01 May 2025; Revised: 17 December 2025; Accepted: 28 January 2026

Abstract: Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, remains a major health challenge, particularly in low income and middle-income countries. This study focuses on the Western province of Sri Lanka, which has high population density, trade, and high human mobility, all contributing to TB transmission. This research aims to strengthen TB control strategies through several key objectives. The TB contact network model was developed and analyzed using graph theoretical measurements such as degree centrality, closeness centrality and betweenness centrality, to identify influential nodes, and structural metrics such as modularity and clustering coefficient to understand transmission patterns. The TOPSIS (technique for order preference by similarity to ideal solution) method was applied to rank nodes based on their potential to spread TB, improving targeted intervention strategies. The novelty of this research is to apply graph theoretical measurements together with the TOPSIS method to identify the highly influential nodes and control the spread of disease by isolating those nodes from the community. Epidemic data extracted from the National Programme for Tuberculosis Control and Chest Diseases (NPTCCD) annual reports (2012–2023) revealed higher projected TB incidences among males and individuals belong to the age group 45–54. The synthetic dataset constructed using percentage-based forecasting data underwent 5-fold cross-validation, and the random forest model achieved an accuracy of 81.5%, indicating strong reliability of the forecasting model. Network analysis indicated the sparse local connections (clustering coefficient: 0.002) and strong community structures (modularity: 0.856). Comparative analysis before and after applying the TOPSIS method demonstrated the improved identification of influential spreaders. Sensitivity analysis showed that the model remained stable under small variations in data and weights, confirming the robustness of the

findings. The outcome of this research provides guidance to health authorities for allocating resources, prioritizing high risk groups, initiating strengthening programmes, and designing more effective community-based interventions to strengthening TB control strategies in the Western province.

Keywords: Centrality measures, cluster analysis, Western province, TOPSIS method, tuberculosis, contact network.

INTRODUCTION

Tuberculosis, abbreviated TB, is an infectious disease usually caused by *Mycobacterium tuberculosis* (MTB) bacteria. Tuberculosis (TB) is primarily spread through the air when a person with TB coughs, sneezes, or talks, expelling tiny droplets containing the bacteria into the air. If another person inhales these droplets, they can become infected with TB. Tuberculosis may infect any part of the body but most commonly affects the lungs; this is known as pulmonary TB. Extrapulmonary TB occurs when tuberculosis develops outside of the lungs, in organs such as lymph nodes, kidneys, bones, and brains (Kant et al., 2021). The World Health Organization (WHO) estimates 10.7 million persons acquired TB in 2024 worldwide and global death among patients with TB exceeds 1.23 million.

Understanding TB transmission dynamics is essential for developing targeted interventions. Over the past decades, researchers have employed various analytical

* Corresponding author (kkkrperera@kln.ac.lk;  <https://orcid.org/0000-0003-2947-1634>)



This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

methods to study the spread of TB. As a South Asian country, Sri Lanka is also facing substantial challenges in the fight against TB. Sri Lanka has made considerable progress in reducing the burden of TB over the past few decades. In the year 2018, Sri Lanka's WHO estimated new and relapsed TB cases were 64 per 100,000 population. By the end of 2018, the estimated number of TB deaths was 3.8 per 100,000 population (excluding TB/HIV deaths) (Ministry of Health, Sri Lanka, 2022). According to the weekly epidemiological reports of Ministry of Health of Sri Lanka, the incidence of TB in 2023 was 41.6 per 100,000 population (Epidemiology Unit of Ministry of Health, 2023), indicating a significant decline in disease transmission.

However, when analyzing TB cases in the Western province, it is required to pay attention to its geographical location, trade, and commerce due to the dense population and significant human mobility.

Several studies were conducted globally to reduce the transmission of TB using statistical approaches as well as social network analysis.

Molecular epidemiology techniques, such as whole-genome sequencing (WGS), help to identify transmission clusters and track strain movements across regions. Studies in Spain (Abascal et al., 2019) and Ghana (Asare et al., 2020) have demonstrated WGS's effectiveness in detecting outbreaks and guiding surveillance efforts. Contact tracing remains a fundamental approach in TB control, identifying individuals exposed to infected patients. (Baluku et al., 2022) have assessed its effectiveness by using logistic regression and Begun et al. (2013) used modeling techniques to optimize detection strategies.

Epidemiological modeling applies mathematical models, including SIR (Side, 2015), SEIR (Side et al., 2017) and agent-based models (Tuite et al., 2017), to simulate TB transmission dynamics and predict disease spread. These models provide insights into infection rates and control strategies. Spatial analysis using GIS (Munch et al., 2003) helps to map TB cases, revealing high-transmission areas. Research in South Africa demonstrated the utility of spatial epidemiology in identifying risk factors and transmission hotspots. Statistical analysis methods, including regression and time series analysis, help assess TB trends and risk factors. Various studies have applied survival models (Burusie et al., 2023) hidden Markov models (Rafei et al., 2012), and Box-Jenkins forecasting (Zheng et al., 2020) to improve TB surveillance. Network analysis examines social interactions that contribute to TB transmission.

Studies using social network analysis (SNA) (Pando et al., 2023) highlighted the role of high-risk communities and provided insights for targeted interventions. When considering the previous studies done in Sri Lanka, most TB studies rely on statistical analysis, with limited adoption of machine learning techniques. Some research has used SPSS for epidemiological assessments, while others have explored deep learning for TB detection via chest X-rays.

The current study is limited to the Western province of Sri Lanka, but can be extended for the whole country as a future work. By leveraging network-based approaches, the study aims to provide a deeper understanding of TB transmission patterns and influential spreaders in this region.

Additionally, the study incorporates the technique for order of preference by similarity to ideal solution (TOPSIS) method with a novel formula proposed to control COVID-19 transmission by isolation of influential nodes (Chaharborj et al., 2022). Rather than treating all contacts equally, decision makers can monitor and prioritize preventive treatment for individuals with the highest transmission potential.

According to the existing literature on TB analysis, social network analysis has been used to explore contact structures and identify high-risk communities, while statistical approaches have been widely used to analyze risk factors, temporal trends, and survival outcomes. Although community detection techniques have been applied in some studies, node-based analysis within detected communities, especially the identification and isolation of influential spreaders, has not been adequately investigated in TB transmission studies. To address this gap, and motivated by recent research that successfully employed TOPSIS-based formulations to control COVID-19 transmission through the identification of influential nodes, the present study integrates TB contact network analysis with the TOPSIS decision making criterion method. This combined framework provides a novel approach for identifying influential TB spreaders within contact networks, offering new insights into TB transmission dynamics in Sri Lanka.

MATERIALS AND METHODS

Data collection involved gathering annual TB reports from the National Programme for Tuberculosis Control and Chest Diseases (NPCCD) website, Ministry of Health, Sri Lanka (Ministry of Health, Sri Lanka, 2022), spanning 2012 to 2023. Based on the annual number of patients

reported during 2012–2023, separate autoregressive integrated moving average (ARIMA) forecasting models were developed using R software to predict the expected TB cases for the year 2025 by gender, age groups, and TB type. The ARIMA forecasting approach assumes that the time series becomes stationary after appropriate differencing, and that the model residuals are independent, normally distributed, and constant variance, and each subgroup model was validated independently using residual diagnostics and forecast accuracy metrics to ensure reliable predictions.

A synthetic dataset of 200 patients was generated based on the forecasted TB cases for 2025, including patient ID, age, gender, contact ID, TB type, and relationship type. The dataset was validated using a Random Forest classifier with 5-fold cross-validation to ensure its predictive stability.

Graph theoretical principles, including centrality measures (betweenness, closeness, degree) and TOPSIS-based node rankings, were applied to analyze TB transmission dynamics. Sensitivity analysis was conducted by perturbing criterion weights and input data to evaluate the robustness of the rankings and disease spread predictions. In this study, graph variants called centrality measures are used to determine the most influential nodes in a given network. Identifying such nodes is important to design resilient networks and identify targets for disease control in epidemic networks. The following gives a brief explanation about the graph theoretical concepts used in this study.

Graph theoretical concepts

Definition 1: Graph

The TB contact network is modelled as a directed graph, where nodes represent patients and directed edges indicate contact or transmission pathways.

Definition 2: Shortest path

The shortest path between two nodes is the minimum number of edges connecting them, representing the most efficient potential route of TB transmission.

Definition 3: Degree

The degree of a node is the number of edges connected to it; in directed graphs, this includes in-degree and out-degree.

Definition 4: Distance

The distance between two nodes is defined as the length of the shortest path separating them in the network.

Definition 5: Betweenness centrality (Wasserman, 1994)

The betweenness centrality (BC) measures the importance of a node in a graph based on its ability to act as a bridge along the shortest paths between pairs of other nodes.

The formula for the betweenness centrality of a node u in a graph is given by,

$g(u) = \sum_{s \neq u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}}$, where σ_{st} is the total number of shortest paths from node s to t , and $\sigma_{st}(u)$ is the number of those paths passing through the node u . Individuals with high betweenness centrality may not have highest number of contacts. However, they can significantly influence disease diffusion due to the prolonged infectious period of the disease.

Definition 6: Closeness centrality (CC) (Rochat, 2009)

The closeness centrality (CC) measures how close a node is to all other nodes in the network on average. Nodes with high closeness centrality can reach other nodes quickly.

The formula for the closeness centrality of a node v in a graph is given by,

$C_{closeness}(v) = \frac{1}{\sum_{u \neq v} dist(u, v)}$, where $dist(u, v)$ is the shortest path distance between nodes u and v .

Definition 7: The degree centrality (DC) (Chaharborj et al., 2022)

The degree centrality (DC) measure is used in network analysis to quantify the number of connections or ties a node has in a network. Nodes with high degree centrality have more close or prolonged interactions, increasing both their risk of acquiring TB and their potential to transmit it to others.

The degree centrality (DC) of node i , denoted as $C_d(i)$, is defined as $C_d(i) = \frac{1}{N} \sum_{j=1}^N x_{ij}$, where, N is the total number of nodes, i is the focus node, j represents all other nodes, x_{ij} represents the connection between node i and node j , and $x_{ij} = \begin{cases} 1, & i \text{ is connected to } j \\ 0, & \text{otherwise} \end{cases}$.

Definition 8: The Clustering coefficient (Watts & Strogatz, 1998)

The Clustering coefficient is a measure of the tendency of nodes to cluster together, while the average clustering

coefficient is the average of these measures across. This explains how disease spreads fast inside the small group but slowly in between groups.

The formula for the clustering coefficient $C(v)$ of a node v in a graph is given by

$C(v) = \frac{2T(v)}{\deg(v)(\deg(v)-1)}$, where $T(v)$ is the number of triangles centered at node v , and $\deg(v)$ is the degree of node v .

Definition 9: The modularity (Newman, 2006)

The modularity is a measure that quantifies how well a network can be divided into distinct modules (also called groups, clusters, or communities). High modularity shows a clear strong community.

For a weighted, directed network, the modularity Q is calculated as follows:

$$Q = \frac{1}{W} \sum_{ij} \left(A_{ij} - \frac{k_i^{\text{in}} k_j^{\text{out}}}{W} \right) \delta(c_i, c_j).$$

Where: A_{ij} is the weight of the edge from node i to node j , k_i^{in} is the in-degree of node i , k_j^{out} is the out-degree of the node j , W is the total sum of edge weights in the network, $\delta(c_i, c_j)$ is the Kronecker delta function, where c_i and c_j are the communities to which nodes i and j belong, respectively.

Definition 10: K-fold cross-validation

K-fold cross-validation is a resampling method used to evaluate the predictive performance of a model. Suppose the dataset D contains N observations, and k is the number of folds. The dataset is partitioned into k approximately equal subsets $\{D_1, D_2, \dots, D_k\}$. For each fold $i = 1, \dots, k$, the model is trained on $D \setminus D_i$ and tested on D_i . Let Err_i denote the prediction error on the i -th fold. Then the overall cross-validated error is computed as:

$$CV_k = \frac{1}{k} \sum_{i=1}^k \text{Err}_i$$

This process ensures that each observation is used exactly once for testing, providing a reliable estimate of model performance (Hastie et al., 2009).

Definition 11: The TOPSIS method

This method is used for ranking in social or complex networks by evaluating each alternative based on its distance from two ideal solutions (Chaharborj et al., 2022):

the positive ideal solution (PIS) and the negative ideal solution (NIS). The positive ideal solution represents the best possible values for each criterion, while the negative ideal solution represents the worst.

Let S_i^+ be the distance from the positive ideal solution and S_i^- be the distance from the negative ideal solution. Then closeness coefficient C_i is $\frac{S_i^-}{S_i^+ + S_i^-}$.

The closeness coefficient calculates how close an alternative is to the ideal solution. This value serves as the basis for ranking the alternatives, with higher C_i indicating a more influential alternative.

The criteria used in this study (degree centrality, betweenness centrality, closeness centrality, clustering coefficient, and modularity) were selected based on their relevance to TB transmission dynamics within the contact network. These criteria quantify the influence of a patient (node) on disease spread, reflecting public health priorities by identifying individuals most likely to contribute to TB transmission. Sensitivity analysis was performed by perturbing weights and input values to ensure the robustness of the TOPSIS rankings under varying assumptions.

Definition 12: Disease spread probability in a complex network

The disease spread probability formula proposed by Chaharborj et al. (2022) is for modeling the transmission dynamics in complex networks. Suppose that G is the proposed graph, N is the total number of nodes, and E is the number of influential nodes with top ranking. Then the pandemic spread probability $P(G_i)$ for the graph G_i can be defined as follows:

$$P(G_i) = \frac{N - i + 1}{N} \times \frac{\sum_{k=1}^N \text{Degree}(G_i)}{\sum_{k=1}^N \text{Degree}(G_0)}, i = 1, \dots, E; G_0 = G$$

Definition 13: Sensitivity analysis (weight- and data-based perturbations)

Sensitivity analysis is a method used to evaluate how variations in model inputs affect the resulting outcomes or rankings. Let a decision matrix $R = [r_{ij}]$ $i = 1, \dots, N, j = 1, \dots, m$, represent N alternatives evaluated under m criteria, and let

$$w = \{w_1, w_2, \dots, w_m\}$$

denote the original criterion weights used in the TOPSIS method.

- **Weight-based perturbation:**

To assess sensitivity with respect to criteria weights, each weight is perturbed as

$$w'_j = w_j + \delta_j, \quad j = 1, \dots, m,$$

subject to the normalization condition

$$\sum_{j=1}^m w'_j = 1$$

where δ_j is a small random perturbation. Using the perturbed weight vector

$$w' = \{w'_1, w'_2, \dots, w'_m\}$$

the TOPSIS procedure is repeated to compute the perturbed closeness coefficient

$$C_i^w = \frac{S_i^-(w')}{S_i^+(w') + S_i^-(w')}, i = 1, \dots, N$$

where $S_i^+(w')$ and $S_i^-(w')$ denote the distances of alternative i from the positive and negative ideal solutions, respectively, calculated using the perturbed weights.

The rankings obtained from C_i^w are then compared with the original TOPSIS rankings to evaluate the stability and robustness of the decision outcomes.

- **Data-based perturbation:**

$$r'_{ij} = r_{ij}(1 + \epsilon_{ij}), \quad \epsilon_{ij} \sim U(-\alpha, \alpha)$$

where ϵ_{ij} is a small random noise added to each criterion value and α is the maximum perturbation magnitude. The perturbed decision matrix $R' = [r'_{ij}]$ is then used to recompute the TOPSIS ranking through the data-perturbed closeness coefficient C_i^d , where C_i^d represents the closeness coefficient obtained after data (decision matrix) perturbation

This process allows assessment of the robustness of TOPSIS rankings and disease spread predictions under small variations in criteria weights and input data, ensuring the stability of the model outcomes (Saltelli et al., 2008; Behzadian et al., 2012; Demir & Arslan, 2022).

The contact network in Figure 1 is a visualization in Gephi that consists of 200 nodes representing patients and 199 edges representing contacts between them. Node

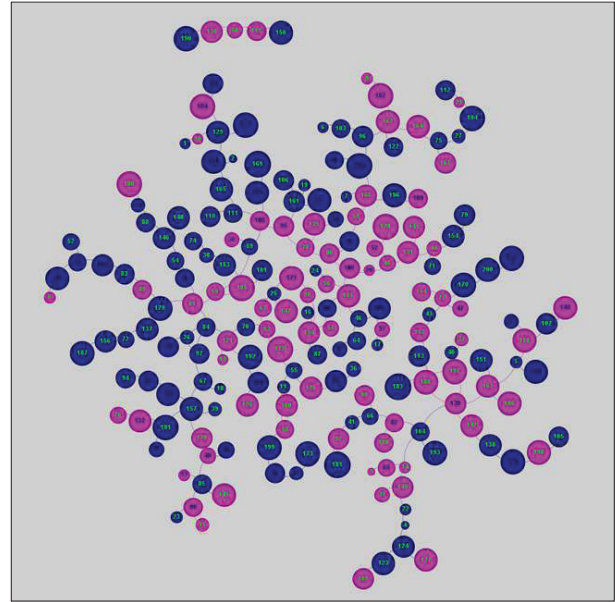
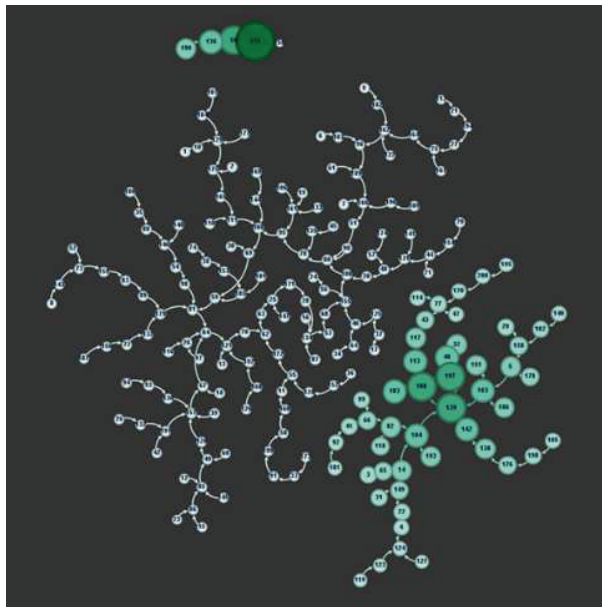


Figure 1: Contact network showing relationship between patients and contacts

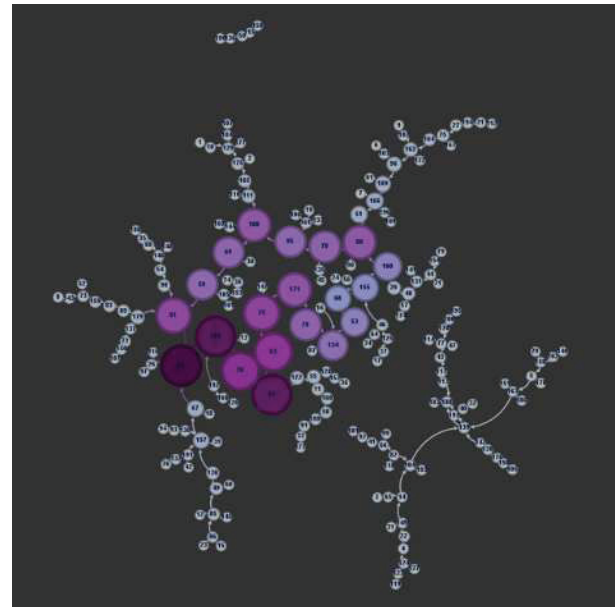
size is used to visualize patient age, with larger nodes indicating older patients and smaller nodes indicating 19 younger patients, providing a clear representation of the age distribution within the network. Gender is visualized using node color, with blue nodes representing male patients and pink nodes representing female patients, allowing for easy identification of the gender distribution in the network. TB type is represented by node label color, with green labels indicating PTB patients and black labels indicating extrapulmonary tuberculosis (EPTB) patients, facilitating the differentiation between different types of TB patients within the network. The visualization reveals interesting patterns, such as higher connectivity among older patients and a balanced distribution of male and female patients in the network

Centrality measures, including closeness centrality and betweenness centrality, were utilized to identify the most influential patients (nodes) in the network. Network visualization of centrality measures is shown in Figure 2.

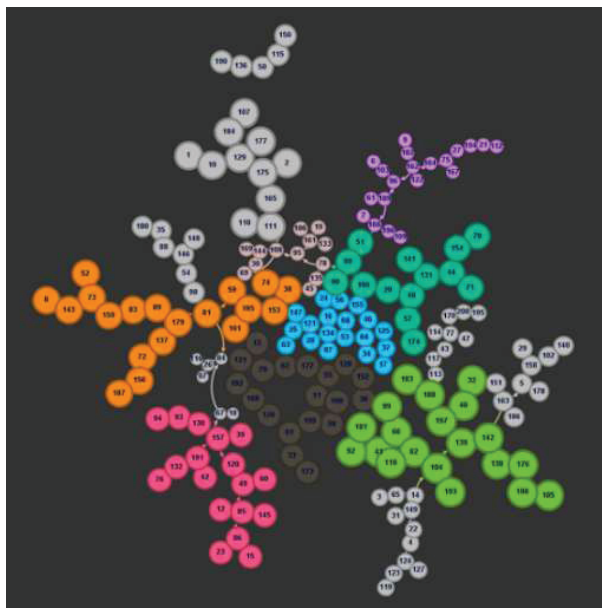
Cluster analysis was carried out to detect densely connected subgroups within the network as shown in Figure 3, and the calculation of average cluster coefficient and modularity provided insights into the network's structural characteristics.



(a) closeness centrality values



(b) betweenness centrality values

Figure 2: Network visualization of highlighting centrality measures**Figure 3:** Network visualization illustrating clustering within the network

The TOPSIS (technique for order of preference by similarity to ideal solution) method was utilized to rank and prioritize nodes based on their likelihood of spreading TB, incorporating multiple criteria for a comprehensive evaluation. The network with fifteen node ranks are highlighted in red colour and represented in Figure 4.

RESULTS AND DISCUSSION

According to the statistical forecasting results, out of a total of 3666 patients, 2439 (67%) were male, and 1227 (33%) were female. Given the lack of real-time data resources, ARIMA models were used to predict the expected number of male and female TB cases for 2025. The forecasted values indicate that male cases are expected to remain higher than female cases, suggesting a persistent gender disparity in TB distribution. Model validation metrics showed RMSE (root mean square error) = 209 for male cases and 127 for female cases, MAE (mean absolute error) = 155 and 94, and MAPE (mean absolute percentage error) = 7.4% and 7.8%, respectively, indicating that the forecasts closely matched historical data.

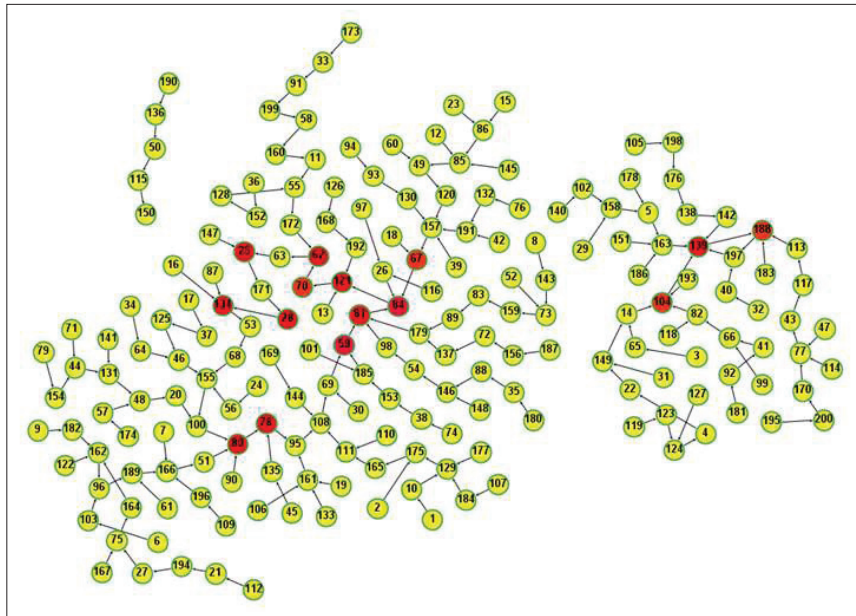


Figure 4: The network visualization of contact network with node ranking

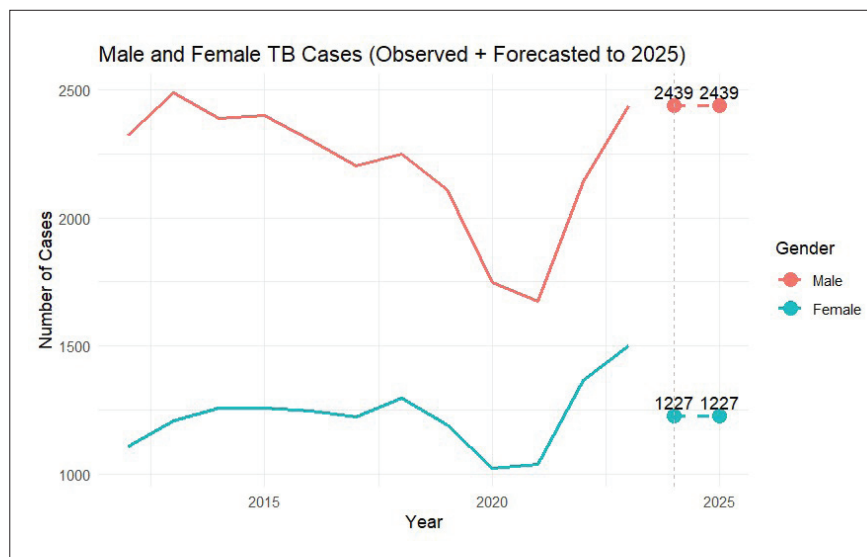


Figure 5: Trend plot of male and female TB cases (2012 – 2023) with forecasts for 2024 - 2025

The age distribution of patients was as follows: 0–14 years: 81 (2%), 15–24 years: 352 (10%), 25–34 years: 387 (11%), 35–44 years: 548 (15%), 45–54 years: 748 (20%), 55–64 years: 708 (19%), 65–74 years: 666 (18%) and 75+ years: 166 (5%). Figure 6 illustrates that the majority of forecasted TB cases in 2025 are expected

to be in the 45–54 age group, accounting for 21% of the overall forecasted cases. Validation of each age-group ARIMA model showed satisfactory predictive performance, with RMSE between 17 and 37 and MAPE between 5.8% and 18.1%, indicating that the forecasts closely matched historical data and are reliable.

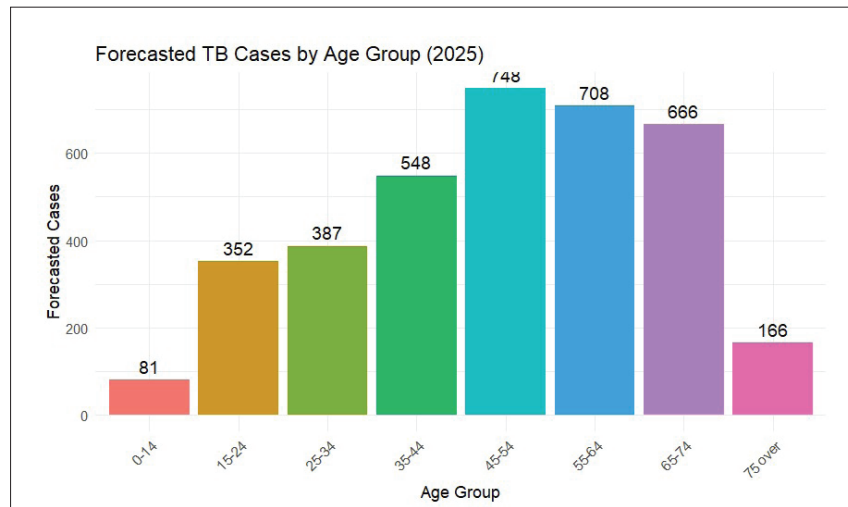


Figure 6: Forecasted TB cases by age group for 2025.

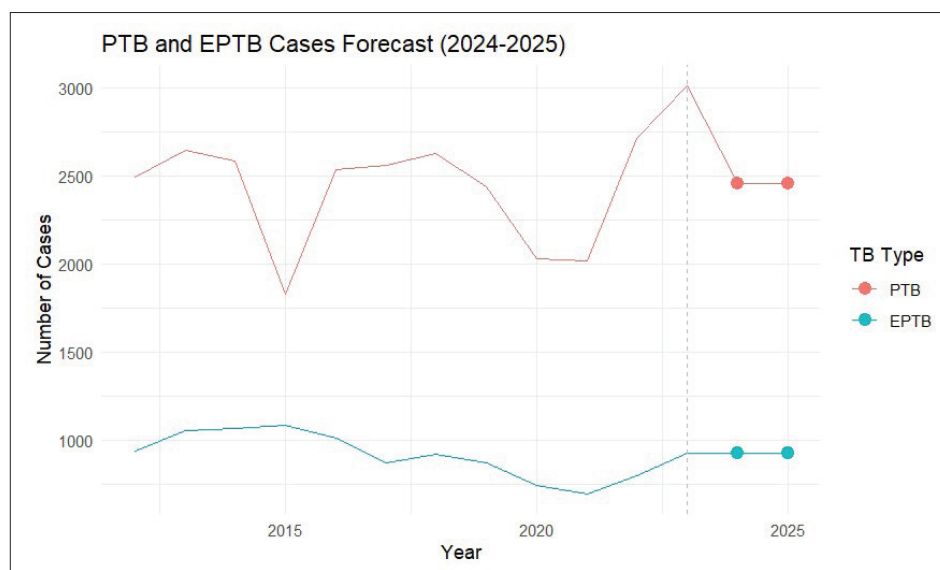


Figure 7: Trend plot of pulmonary tuberculosis (PTB) and extrapulmonary tuberculosis (EPTB) cases with forecast for 2025.

Regarding TB types, 2457 (73%) had pulmonary TB (PTB) and 926 (27%) had extrapulmonary TB (EPTB). Forecasted values for trends of PTB and EPTB are shown in Figure 7. Both PTB and EPTB cases fluctuated over the years, with PTB cases consistently higher than EPTB. Forecast validation showed RMSE values of 322 for PTB and 86 for EPTB, with MAPE values of 11.1% and 8.3%, indicating that the forecasts closely matched

historical data. PTB cases are expected to remain the majority of TB cases in 2025.

Based on the forecasting data, a comprehensive data table was constructed using R software with a 200-patient sample representing the forecasting data population. A Random Forest model was applied to classify TB type (PTB vs EPTB) using three predictors. The model was

tuned with 5-fold cross-validation, and the optimal mtry = 2 (number of predictors randomly selected at each split) achieved an accuracy of 81.5%, indicating reliable predictive performance for distinguishing between PTB and EPTB cases.

We visualize high centrality values using larger nodes and darker colours. Nodes with high closeness values are in central positions within the network, allowing them to reach other nodes quickly. Nodes with high betweenness centrality, marked with special visual cues, act as important links in the network, connecting different groups of nodes.

The average clustering coefficient of 0.002 in the TB network for the Western Province indicates a low level of local clustering among TB cases, suggesting relatively few clusters of interconnected cases. In contrast, the modularity value of 0.856 signifies a high level of community structure within the network, indicating that TB cases are more densely connected within specific communities or groups. These findings reveal that TB cases are not highly connected locally but tend to form distinct, well-defined groups or communities.

Table 1: Node ranking based on relative closeness to the ideal solution

Nodes	S_i^+	S_i^-	C_i	Ranking
1	0.0252	0.0269	0.5158	14
11	0.0246	0.0283	0.5357	2
14	0.0247	0.0276	0.5276	7
15	0.0247	0.0274	0.5261	9
17	0.0248	0.0276	0.526	10
31	0.0249	0.0266	0.5164	13
42	0.0245	0.0283	0.5353	4
49	0.0246	0.0279	0.531	6
53	0.0249	0.0274	0.5247	11
56	0.0248	0.0272	0.5232	12
61	0.0246	0.0284	0.5354	3
82	0.0254	0.0267	0.5127	15
103	0.0245	0.0284	0.5365	1
170	0.0248	0.0281	0.5312	5
180	0.0247	0.0276	0.5272	8

Here, S_i^+ = The distance from the positive ideal solution, S_i^- = The distance from the positive ideal solution and C_i = Closeness coefficient.

Table 2: Disease spread probability in the network with varying numbers of nodes removed.

Number of removed nodes	Disease spread probability
1	0.9900
2	0.9751
3	0.9554
4	0.9407
5	0.9163
6	0.8970
7	0.8827
8	0.8733
9	0.8688
10	0.8500
11	0.8408
12	0.8221
13	0.8037
14	0.7948
15	0.7859

The results of node ranking based on relative closeness to the ideal solution are presented in Table 1, identifying top-ranking individuals crucial for understanding TB transmission pathways within the network. The highest C_i value 0.5365, is ranked first among the 200 nodes, indicating its significant influence on the network. The disease spread probability decreases as more top-ranking nodes are removed, dropping from 0.99 with one node removed to 0.7859 with fifteen nodes removed, as shown in Table 2.

The results reveal key trends in TB transmission in the Western province of Sri Lanka, indicating a shift that requires gender-specific interventions. Most TB cases forecasted for 2025 are expected in the 45-54 age group, emphasizing the need to target this age group for effective TB control. Pulmonary TB (PTB) remains more prevalent than extrapulmonary TB (EPTB), but both types have decreased since 2020, suggesting improvements in TB control measures.

Figure 8 shows how the chance of the disease spreading changes under three situations: original, data perturbed, and weight perturbed, as more nodes are removed. It shows that the model is strong and reliable even if the patient data or the criteria weights are slightly changed. In all three cases, the chance of disease spreading goes down as more important nodes are removed, meaning that isolating key patients or reducing their contacts can

help control the spread. The three lines are very similar, which means the results are consistent. Small differences appear: the data perturbed line is a little lower, and the

weight perturbed line is sometimes a little higher, but these differences are very small and do not change the main trend.

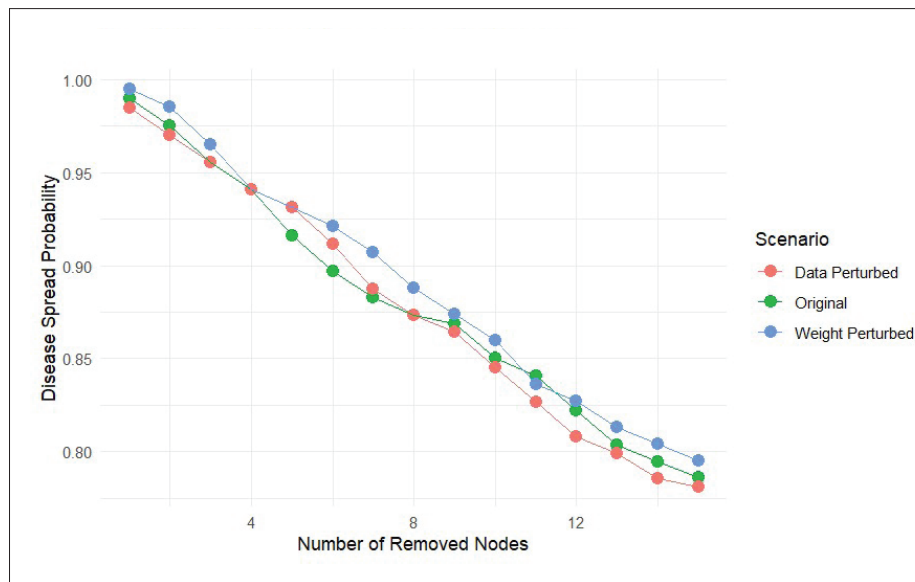


Figure 8: Sensitivity analysis of disease spread probability in the contact network

The data table and contact network visualizations highlight the demographics and relationships among patients, with older patients showing high connectivity and a balanced gender distribution. Close friends and household connections play a significant role in TB transmission, suggesting that interventions should target these relationships to mitigate spread. Network centrality measures provide further insights. Higher centrality values indicate key nodes that are more connected and influential; closeness centrality reflects how quickly a node can reach others, while betweenness centrality identifies key mediators of information flow. The low clustering coefficient and high modularity value suggest sparse local connections but strong community structures.

Identifying top-ranking nodes through the TOPSIS approach provides insights into potential TB transmission pathways. Removing these influential nodes effectively reduces TB spread, underscoring the importance of targeting key individuals and communities to improve TB control and treatment outcomes in the Western province of Sri Lanka. Overall, the sensitivity analysis confirms that the model is both responsive to interventions and

robust to small changes in inputs, supporting its reliability for guiding decision-making in intervention strategies.

This study focused only on the Western Province, which has high population density. Human mobility from outside the province could affect TB transmission but was not included in the analysis. Patient data collection was challenging because real healthcare data is not easily accessible, and public open-source data from websites or other sources are limited. Some cases may be underreported or reported late, which could affect data accuracy. Differences in contact patterns between individuals and unmeasured socio-economic or environmental factors could also influence TB spread.

Despite these limitations, the study provides guidance for health authorities. By discovering strong community structures, health authorities can prioritize influential spreaders for targeted screening and isolation, allocate limited resources more efficiently, and conduct community-wise awareness programs rather than applying uniform interventions. Overall, more effective age and gender specific interventions, as well as workplace or trade-based screening programs, can be planned.

CONCLUSION

This study integrated statistical forecasting, network analysis using graph theory, and the TOPSIS method to investigate TB transmission dynamics in Western province, Sri Lanka. Our study focused on analyzing TB spread in this region, utilizing data-driven methods to identify key trends and influential spreaders. Our analysis reveals that males have a higher percentage of TB cases in 2025. The 45-54 age group emerges as a significant demographic for TB incidence, with pulmonary TB (PTB) forecasted to constitute 73% of cases in 2025. Network analysis provided insights into disease transmission pathways, highlighting community structures and identifying influential nodes crucial for targeted interventions. The low average clustering coefficient suggests sparse local connections among TB cases, while the high modularity value indicates strong community structures where TB cases are more densely connected within specific groups than across the entire network. By focusing on key nodes and community structures, interventions can effectively mitigate TB spread. Given the lack of comprehensive data beyond 2025, our forecasting focuses on this year, ensuring that predictions are based on the most recent and reliable information available. Despite focusing on 2025, the methodological framework and insights from the study offer valuable guidance for future research and public health strategies. Additionally, the sensitivity analysis confirmed that the model's results remain stable under small data and weight perturbations, further strengthening the reliability of the findings. The findings of this research are helpful to health authorities to take quick decisions regards resource allocations, identify high risk group, and isolate influential patients, in order to control the spread of a disease.

REFERENCES

- Abascal, E., Pérez-Lago, L., Martínez-Lirola, M., Chiner-Oms, Á., Herranz, M., Chaoui, I., *et al.* (2019). Whole genome sequencing-based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations. *Eurosurveillance*, 1800005. doi:https://doi.org/10.2807/1560-7917.es.2019.24.4.1800005
- Asare, P., Otchere, I. D., Bedeley, E., Brites, D., Loiseau, C., Baddoo, N. A., *et al.* (2020). Whole genome sequencing and spatial analysis identifies. *Frontiers in medicine*, 161. doi:https://doi.org/10.3389/fmed.2020.00161
- Baluku J.B., Nabwana M., Winters M. & Bongomin F. (2022). Tuberculosis contact tracing yield and associated factors in Uganda. *BMC Pulmonary Medicine*, 22, 1-7. doi:https://doi.org/10.1186/s12890-022-01860-z
- Begun M., Netall A.T., Marks G.B. & Wood J.G. (2013). Contact tracing of tuberculosis: a systematic review of transmission modelling studies. *Plos one*. doi:https://doi.org/10.1371/journal.pone.0072470
- Behzadian, M., Otaghsara, S. K., Yazdani, M., & Ignatius, J. (2012). A state-of-the-art survey of TOPSIS applications. *Expert Systems with Applications*, 39, 13051–13069. doi:10.1016/j.eswa.2012.05.056
- Burusie, A., Enquesilassie, F., Salazar-Austin, N., & Addissie, A. (2023). Epidemiology of childhood tuberculosis and predictors of death among children on tuberculosis treatment in central Ethiopia: an extended Cox model challenged survival analysis. *BMC public health* 23.1, 23, 1287. doi:https://doi.org/10.1186/s12889-023-16183-9
- Chaharborj S.S., Nabi K.N., Feng K.L. & Phang P.S. (2022). Controlling COVID-19 transmission with isolation of influential nodes. *Chaos, Solitons & Fractals*, 159, 112035. doi:https://doi.org/10.1016/j.chaos.2022.112035
- Demir, G., & Arslan, R. . (2022). Sensitivity Analysis in Multi-Criterion Decision-Making Problems. *Ankara Hacı Bayram Veli University Journal of Economics and Administrative Sciences*, 24(3), 1025–1056.
- Epidemiology Unit of Ministry of Health. (2023). WEEKLY EPIDEMIOLOGICAL REPORT. 59(9), 1-5.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.)*. Springer.
- Kant, S., Mitta, V., Mavi, A. K., Chaturvedi, R., & Singh, P. (2021). Tuberculosis: An overview and review of literature. *Acta Pharmacologica Sinica*, 64-68.
- Ministry of Health, Sri Lanka . (2022). *Tuberculosis: the best public health practice*. Retrieved from https://www.epid.gov.lk/storage/post/pdfs/vol_49_no_37-english_1.pdf
- Munch, Z., Van Lill, S. W. P., Booysen, C. N., Zietsman, H. L., Enarson, D. A., & Beyers, N. (2003). Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *International journal of tuberculosis and lung disease* 7.3, 7, 271-277.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103, 8577--8582. doi:https://doi.org/10.1073/pnas.0601602103
- Pando, C., Hazel, A., Tsang, L. Y., Razafindrina, K., Andriamiadanarivo, A., Rabetombosoa, R. M., *et al.* (2023). A social network analysis model approach to understand tuberculosis transmission in remote rural Madagascar. *BMC Public Health*, 23, 1511. doi:https://doi.org/10.1186/s12889-023-16425-w
- Rafei, A., Pasha, E., & Orak, R. J. (2012). Tuberculosis surveillance using a hidden Markov model. *Iranian Journal of Public Health* 41.10, 41, 87.
- Rochat, Y. (2009). Closeness Centrality Extended To Unconnected Graphs :The Harmonic Centrality Index. In *Asna* (pp. 26-28).
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., *et al.* (2008). *Global Sensitivity Analysis: The*

- Primer*. Chichester, UK: John Wiley & Sons.
- Side, S. (2015). A susceptible-infected-recovered model and simulation for transmission of tuberculosis. *Advanced Science Letters* 21.2, 21, 137-139. doi:<https://doi.org/10.1166/asl.2015.5840>
- Side, S., Mulbar, U., Sidjara, S., & Sanusi, W. (2017). A SEIR model for transmission of tuberculosis. *AIP conference proceedings*, 1830. doi:<https://doi.org/10.1063/1.4980867>
- Tuberculosis*. (n.d.). Retrieved from NPTCCD: <https://www.nptccd.health.gov.lk/>
- Tuite, A. R., Gallant, V., Randell, E., Bourgeois, A. C., & Greer, A. L. (2017). Stochastic agent-based modeling of tuberculosis in Canadian Indigenous communities. *BMC public health*, 1-12. doi:<https://doi.org/10.1186/s12889-016-3996-7>
- Wasserman, S. (1994). Social Network Analysis: Methods and Applications. *The Press Syndicate of the University of Cambridge*. doi:<https://doi.org/10.1017/CBO9780511815478>
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393, 440--442. doi:<https://doi.org/10.1038/30918>
- World Health Organization*. (2023). Retrieved from Wikipedia: <https://www.overleaf.com/project/65d6b43afac6df0c68dc8503#:~:text=https%3A//www.who..>
- Zheng, Y., Zhang, L., Wang, L., & Rifhat, R. (2020). Statistical methods for predicting tuberculosis incidence based on data from Guangxi, China. *BMC infectious diseases*, 20, 1-8. doi:<https://doi.org/10.1186/s12879-020-05033-3>