Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

Venue-Specific Probabilistic Modeling of High Run Chases in T20I Cricket: A Bayesian Network Approach

Syed Asghar Ali Shah (Corresponding Author)
Department of Statistics, University of Peshawar, Pakistan
Email:syedasghar82@uop.edu.pk

Qamruz Zaman
Department of Statistics, University of Peshawar, Pakistan
Email: cricsportsresearchgroup@gmail.com

Ubaid-Us-Salam Qazi
Department of Statistics, Govt. Postgraduate College, Nowshera, KP
Email: qus28jan@gmail.com

Amjad Ali
Principal Research Officer, Cereal Crops Research Institute Pirsabak, Nowshera
Email: srobreeding@gmail.com

Touheed Iqbal
Senior Research officer, Agriculture Department (Research Wing) Peshawar
Email: touheed.ik@gmail.com

Shahid Ali
Research Officer, Cereal Crops Research Institute, Pirsabak,Nowshera
Email: hi2shahid@gmail.com

Muhammad Waqas
Cereal Crop Research Institute, Pirsabak, Nowshera
Email: m.waqas.ent@gmail.com

Abstract
This study investigates the impact of venue-specific factors on high run chases in T20 International (T20I) cricket using Bayesian Networks (BNs) derived from Gaussian Graphical Models (GGMs). Analyzing data from 458 high-scoring matches between 2005 and 2024, the study incorporates variables such as toss outcomes, pitch conditions, team rankings, and match results to construct probabilistic models tailored to home, neutral, and away venues. Regularization through Graphical Lasso ensured sparsity, yielding interpretable networks with optimal complexity. Results reveal that venue conditions significantly influence dependency structures, with away venues exhibiting denser interdependencies and requiring greater adaptability. Critical factors identified include Toss Outcome (TO), Toss Decision (TD), and Pitch Conditions (PC), while Result (R) consistently emerged as the most influential variable across all venues. Networks with 16 edges provided the best balance of fit and simplicity, validated through posterior probabilities. These findings highlight the importance of venue-specific strategies for optimizing high run chases in T20I cricket. The study advances the

application of probabilistic modeling in cricket analytics, offering actionable insights for teams and decision-makers. Future research could integrate additional factors, such as player metrics and weather conditions, to refine predictive models further and expand their applicability across different cricket formats.
Key Words: Cricket Analytics, Bayesian Network, Gaussian Graphical Models, Probabilistic Modeling, Venue-Specific factors, Sparsity

Introduction

The T20 format of cricket has transformed the sport, introducing an exciting fast-paced dynamic where high-scoring matches are common. Among these, successful high run chases demand exceptional strategic planning and execution. Such outcomes rely on a complex interplay of factors, including pitch conditions, toss decisions, match settings, and team performance (Shah *et al.*). Understanding these dependencies has become increasingly important for analysts and teams. Over the years, researchers have explored deterministic and probabilistic approaches to model these dynamics and extract actionable insights. For instance, Isaacs and Finch (2021) emphasized the need for advanced statistical models to capture the intricate factors influencing match outcomes, especially in high-scoring games.

Probabilistic models, particularly Bayesian Networks (BNs), have emerged as a robust framework in sports analytics due to their ability to represent uncertainties and dependencies among variables. Representing relationships as directed acyclic graphs, BNs provide a way to predict outcomes and infer the influence of variables like toss outcomes, batting orders, and rankings on match results. Pearl (1988) introduced Bayesian reasoning as a structured approach for analyzing complex systems, which has since found applications across various domains, including cricket. More recently, Samad (2019) demonstrated the use of probabilistic models in cricket analytics, showing how BNs can identify critical factors that provide run-scoring advantages, offering valuable insights for tactical decision-making.

Gaussian Graphical Models (GGMs) complement Bayesian Networks by identifying conditional dependencies through the precision matrix, which represents the inverse of the covariance matrix. GGMs are effective in high-dimensional datasets, as they focus on reducing complexity by eliminating weaker relationships through regularization techniques like Graphical Lasso (Friedman et al., 2008). This sparsity allows for clearer, interpretable networks that highlight meaningful connections between variables. GGMs are particularly relevant for T20 cricket, where they can pinpoint critical dependencies among factors such as toss outcomes, pitch conditions, and team rankings, providing the structural foundation for Bayesian Network models.

The impact of venue conditions whether home, neutral, or away on match outcomes is a critical area of analysis in cricket. Pollard (2008) and Clarke and Norman (2009) highlighted that home advantage often stems from familiarity with pitch conditions, crowd support, and reduced fatigue due to less travel. Neutral venues, on the other hand, minimize these advantages, emphasizing strategic adaptability, while away venues present significant challenges due to unfamiliar conditions and external pressures. These venue-specific dynamics are crucial in understanding how factors like toss decisions and team rankings interact to influence match results, particularly in the context of high run chases. This study explores these relationships, leveraging probabilistic methods to analyze venue-specific variations.

While cricket analytics has advanced significantly, there is limited research on modeling high run chases using probabilistic approaches tailored to venue-specific dynamics. Although existing studies have examined individual factors, few have integrated Gaussian Graphical Models with Bayesian Networks to offer a comprehensive probabilistic framework. This study addresses this gap by analyzing data from 458 high-scoring T20I matches played between 2005 and 2024. By incorporating GGM-based Bayesian Networks, it aims to identify the most influential factors affecting match outcomes and evaluate how these relationships vary across different venues.

This research makes three significant contributions. Firstly, it applies a novel GGM-based approach to construct Bayesian Networks, effectively capturing the probabilistic dependencies between factors. Secondly, it offers a detailed venue-wise analysis, revealing how conditions like toss outcomes, pitch types, and team rankings influence high run chases. Lastly, the study advances probabilistic modeling in cricket analytics, providing a foundation for future research into predictive modeling and decision-making in other sports. These findings have practical implications for improving strategies, particularly in international competitions played across diverse venues, highlighting the role of integrated probabilistic models in advancing sports analytics.

Literature Review

In recent years, the use of probabilistic approaches in sports analytics has gained momentum, particularly in cricket. Bayesian Networks (BNs) are a prominent method for representing and analyzing dependencies between variables in a structured, interpretable manner. Heckerman and Geiger (1995) demonstrated how BNs could effectively model probabilistic relationships, making them invaluable in scenarios requiring uncertainty quantification and decision-making. Their relevance in cricket analytics has grown as factors like toss outcomes, team strategies, and batting orders significantly impact match outcomes. Mohan, Pearl, and Tian (2013) further highlighted the role of graphical models in causal reasoning, underscoring their capacity to handle complex interdependencies.

The integration of probabilistic models with machine learning has also advanced cricket analytics. López-Pintado et al. (2021) introduced hybrid models that combine Bayesian reasoning with regression-based methods to predict player and team performance. These models enhance interpretability while maintaining predictive accuracy, making them well-suited for high-stakes formats like T20 cricket. Similarly, Pereira et al. (2015) demonstrated the application of Bayesian classifiers in analyzing cricket matches, showing their ability to incorporate domain-specific knowledge to improve predictive capabilities.

Gaussian Graphical Models (GGMs) have become a cornerstone in the development of Bayesian Networks, particularly when dealing with high-dimensional data. GGMs use the precision matrix, derived as the inverse of the covariance matrix, to identify conditional dependencies among variables. Lauritzen (1996) emphasized their utility in constructing sparse networks, which are easier to interpret while preserving critical relationships. Regularization

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

techniques like Graphical Lasso, introduced by Friedman, Hastie, and Tibshirani (2008), have been pivotal in ensuring that weaker dependencies are excluded, enabling more focused and interpretable analyses. These methods are especially valuable in cricket, where numerous interacting factors, such as pitch conditions, match settings, and rankings, require systematic exploration.

Venue-specific dynamics significantly influence match outcomes, as playing conditions vary considerably between home, neutral, and away venues. Research by Pollard (2008) and Clarke and Norman (2009) showed that home advantage often results from factors such as pitch familiarity, crowd support, and reduced travel fatigue. Neutral venues, by contrast, equalize conditions, emphasizing adaptability and strategic planning, while away matches present challenges due to unfamiliar environments and external pressures. Narayanan and Rajagopal (2020) extended this analysis, exploring how venue conditions impact team performance and strategy. Ahmed et al. (2022) highlighted the importance of pitch conditions as a critical factor in batting performance, further emphasizing the need for venue-specific analyses.

The use of probabilistic models to predict and analyze match outcomes has grown substantially, with applications extending to causal inference. Peters et al. (2014) introduced methods for using invariant predictions in sports analytics, allowing analysts to estimate the effects of specific strategies, such as batting order changes or field placements, on match results. These causal models are particularly relevant in T20 cricket, where real-time decision-making plays a crucial role in determining outcomes. However, integrating domain-specific constraints with probabilistic models remains a challenge. Smith et al. (2021) proposed an approach that incorporates expert knowledge into statistical models, improving their relevance and interpretability while enhancing decision-making processes in sports analytics.

The literature demonstrates that probabilistic frameworks like Bayesian Networks and GGMs are instrumental in cricket analytics, providing tools to model complex interdependencies and generate actionable insights. These methods continue to advance the field, paving the way for more precise analyses and practical applications in decision-making across sports.

Methodology

The study investigates venue-wise factors influencing high run chases in T20I cricket using Bayesian Networks derived from Gaussian Graphical Models (GGMs).

Data Collection

Data from 458 T20I high scoring matches (2005–2024) were collected including variables such as Venue (home, neutral, away), Pitch Conditions, Match Conditions, Batting Order, Toss Outcome, Toss Decision, Team Rankings, and Result. The detail descriptions of factors with their levels are shown in Table 1 below;

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 2 No. 4 (November) (2024)

Table 1: Description of factors affecting team performance

Factor Name	Status	Level	Level Label
Result/outcome	won, lost	2	1:won 0: lost
Venue(Vn)	Home, away, neutral	3	-1: away 0:neutral 1: home
Match condition(MC)	day, night, day/night	3	-1: night 0: day/nigh 1: day
Toss outcomes(TO)	won, lost	2	0: lost 1: won
Toss decision(TD)	bat, field	2	0: won the toss and field 1: lost the toss and bat
Batting Order	low, medium, upper	3	-1: represents lower order from Player 8 to 10 0: displays middle order from player 4 to 7. 1: signifies upper order of player from position 1 to 3.
Pitch condition (PC)	slow & low, balanced, Batting pitch, bowling pitch	3	-1: indicates slow and low. 0: denotes balanced. 1: show a batting friendly 2: represent bowling friendly
Ranking of Team	weak, average , strong	3	-1: indicate ranking above 10 0: represent ranking >5 and <10 1: display ranking from 1 to 5
Ranking of opponent	weak, average , strong	3	-1: indicate ranking above 10 0: represent ranking >5 and <10 1: display ranking from 1 to 5

Statistical Analysis

GGMs were used to identify conditional dependencies through precision matrices and regularized with Graphical Lasso for sparsity. The undirected GGM structure was converted into a Directed Acyclic Graph (DAG) for Bayesian Network analysis. Venue-specific networks were analyzed for sparsity, edge evidence probabilities, and posterior probabilities, revealing optimal complexity at 16 edges. Centrality measures (Betweenness, Closeness, Strength, Expected Influence) identified key variables like Toss Outcome, Pitch Conditions, and Result across venues. Validation was conducted using posterior probabilities, and network visualizations highlighted dependency strengths. JASP was used for GGM analysis, and R software supported Bayesian Network modeling and visualization.

Gaussian Graphical Model (GGM)

A GGM is an undirected graphical model used to identify conditional dependencies between variables. In this study, the GGM was employed to analyze probabilistic relationships among factors like Toss Outcome, Pitch Conditions,

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

and Result. The precision matrix Θ, derived from the covariance matrix Σ, was used to estimate these dependencies.

$$\Theta = \Sigma^{-1} \dots \dots \dots (1)$$

In a GGM, if θ_{ij} (off-diagonal element of Θ) is zero, variable X_i and X_j are conditionally independent. For example, relationships between Toss Outcome (TO) and Toss Decision (TD) are captured by non-zero elements in Θ. Regularization through Graphical Lasso ensured sparsity, resulting in interpretable networks by penalizing weaker connections.

Directed Acyclic Graph (DAG)

A DAG is a graphical structure where directed edges represent probabilistic influences, and cycles are prohibited. In the venue-wise analysis, the undirected GGM structure was converted into a DAG to model directional dependencies, such as how Toss Outcome (TO) influences Toss Decision (TD) and how these collectively affect the Result (R).
A DAG is a probabilistic representation where each directed edge i→j indicates a dependency. The joint probability distribution is factorized as:

$$\mathrm{P(X_1, X_2, \dots .. X_n)} = \prod_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{P\big(X_i \big| Parents(X_i)\big)} \dots \dots .. (2)$$

For instance, the probability of Result (R) depends on Toss Decision (TD) and Pitch Conditions (PC):

$$\mathrm{P(R) = P(R|TD, PC)} \dots \dots \dots .. (3)$$

DAGs are constructed from the GGM structure to model these relationships.

Sparsity

Sparsity indicates the proportion of non-zero edges (connections) in a network relative to the total possible edges. Networks with high sparsity (fewer edges) are simpler, while lower sparsity reflects denser dependencies.
For a network with "n" nodes, the total possible edges are given by:

$$Total\ Edges = \frac{n(n-1)}{2} \dots \dots .. (4)$$

Sparsity(S) is defined as;

$$S = 1 - \frac{Non - zero\ Edges}{Total\ Edges} \dots \dots .. (5)$$

In this study home venue had 12 edges, yielding;

$$S = 1 - \frac{12}{28} = 0.571$$

The home venue network had 12 edges (sparsity = 0.571), indicating fewer and more selective dependencies, whereas the away venue network had 14 edges (sparsity = 0.500), showing more complex relationships.

Edge Evidence Probability

The edge evidence probability quantifies the likelihood of a connection between two variables, calculated as:

$$P\big(Edge_{ij}\big|Data\big) = \frac{\int P(Data|Edge_{ij}, \Theta)P(\Theta)}{\int P(Data|\,\Theta)\,P(\Theta)} \dots \dots \dots (6)$$

For example, the strong connection between Toss Decision (TD) and Result (R) at away venues had P=1.000.

Posterior Probability

Posterior probability evaluates the likelihood of a network structure ς given data Ɗ is;

$$P(\varsigma|Ɗ) = \frac{(P(Ɗ|\varsigma)\, P(\varsigma))}{P(Ɗ)} \dots \dots (7)$$

Here, $(P(Ɗ \mid \varsigma)$ is the likelihood, $P(\varsigma)$ is the prior probability of the structure, and $P(Ɗ)$is the marginal likelihood. Networks with 16 edges had the highest posterior probability in all venues, balancing complexity and fit.

Centrality Measures

Centrality measures assess the importance of variables within the network:

- **Betweenness**: Highlights how often a variable bridges others (e.g., Toss Outcome (TO) at away venues).

$$Between(X_i) = \sum_{s \neq i \neq t} \frac{\delta_{st}(i)}{\delta_{st}} \dots \dots .(8)$$

- **Closeness:** Measures the ease with which a variable influences others (e.g., Pitch Conditions (PC) at neutral venues).

$$Clossesness(X_i) = \frac{1}{\sum_{j \neq i} d(X_i, X_j)} \dots \dots .(9)$$

  Where $d(X_i, X_j)$ is the shortest path distance between X_i and X_j.
- **Strength**: Captures the direct cumulative impact of a variable (e.g., Result (R) across all venues).

$$Strength(X_i) = \sum_j w_{ij} \dots \dots \dots (10)$$

  where w_{ij} is the weight of the edge between X_i and X_j.
- **Expected Influence**: Reflects the overall effect of variables, showing Result (R), Toss Outcome (TO), and Pitch Conditions (PC) as the most influential factors.

Complexity

It refers to the number of edges in a network. The relationship between network complexity and posterior probability is captured by plotting the number of edges (E) against posterior probability (P):

$$P(E) \propto P\big(D \mid G(E)\big)P\big(G(E)\big)P(E) \dots \dots (11)$$

Moderately complex networks with 16 edges showed peak posterior probabilities, emphasizing the balance between simplicity and explanatory power.

Bayesian Network (BN)

A BN is a directed probabilistic model constructed from the GGM. The BN represents the joint distribution of variables as a DAG, combining structure and parameters:

$$P(X_1, X_2, \dots \dots \dots X_n) = \prod_{i=1}^{n} P(X_i|Parents(X_i) \dots \dots \dots (12)$$

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

For example, the probability of a match result (R) depends on Pitch Conditions (PC) and Toss Decision (TD):

$$P(R) = P(R|PC, TD) \ldots\ldots\ldots (13)$$

Precision Matrix

The precision matrix, the inverse of the covariance matrix, identifies direct relationships between variables while controlling for others. The precision matrix, Θ, identifies conditional independencies. For variables Xi and X_j, the partial correlation coefficient is:

$$\rho_{ij} = -\frac{\theta_{ij}}{\sqrt{\theta_{ii}\theta_{jj}}} \ldots\ldots..(14)$$

Here, θ_{ij} is the off-diagonal element of Θ. A zero value for θij implies X_i and X_j are conditionally independent.

It forms the backbone of the GGM and was critical for understanding conditional dependencies such as the strong influence of Toss Decisions on Results at neutral and away venues.

Regularization

Regularization minimizes over fitting in network estimation. In Graphical Lasso, the optimization problem for the precision matrix Θ is:

$$\widehat{\Theta} = \arg \min_{\Theta} [\text{trace } (\Sigma\Theta - \log \det (\Theta) + \lambda \|\theta\|_1] \ldots\ldots (15)$$

Where λ is the penalty term controlling sparsity. This ensures only significant edges are retained.

Results and Discussion

Table 2 summarizes the Bayesian Network structure derived using the Gaussian Graphical Model (GGM) approach for analyzing venue-wise effects on T20I cricket outcomes, where venue levels are encoded as 1 (home), 0 (neutral), and -1 (away). Each network corresponds to one venue scenario, with 8 nodes representing variables (e.g., Pitch Conditions, Match Conditions, Batting Order, etc.).

Table 2: Venue wise summary of network for factor affecting high run chase

Summary of Network ▼

Network	Number of nodes	Number of non-zero edges	Sparsity
-1	8	14 / 28	0.500
0	8	13 / 28	0.536
1	8	12 / 28	0.571

The total possible connections (edges) between nodes is 28, calculated based on the undirected network structure (n (n−1)/2n (n-1)/2n (n−1)/2 where n=8n = 8n=8). For the home venue (1), the network has 12 non-zero edges with a sparsity of 0.571, indicating fewer connections and a more selective dependency structure. The neutral venue (0) network has 13 edges and a sparsity of 0.536, slightly denser than the home network. In contrast, the away venue (-1) network has the most connections (14 edges) and the lowest sparsity (0.500), suggesting that away matches exhibit the most complex interdependencies among variables. These differences in sparsity reflect how venue impacts the dependency structure

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

of variables, with home conditions potentially simplifying dependencies, while away matches demand a more nuanced interplay among factors. This analysis highlights how the venue influences the probabilistic relationships between variables, aiding in venue-specific decision-making in cricket analytics.

Table 3 presents the Edge Evidence Probability for Bayesian Network analysis using the Gaussian Graphical Model (GGM) approach, broken down by venue conditions: -1 (away), 0 (neutral), and 1 (home). Each cell represents the probability of an edge (dependency) existing between two variables based on the GGM framework. The variables include Batting Order (BO), Match Condition (MC), Toss Outcome (TO), Toss Decision (TD), Pitch Conditions (PC), Ranking of Teams (RoT), Opponent Team Ranking (Ropp), and Result (R). High probabilities (close to 1) indicate a strong dependency between variables, while lower probabilities (closer to 0) suggest weaker or negligible relationships.

Table 3: Venue wise edge evidence probability for factor affecting high run chase

Edge evidence probability table

	-1								0								1							
Variable	BO	MC	TO	TD	PC	RoT	Ropp	R	BO	MC	TO	TD	PC	RoT	Ropp	R	BO	MC	TO	TD	PC	RoT	Ropp	R
BO	0.000	0.060	1.000	1.000	0.030	0.140	0.160	1.000	0.000	0.090	1.000	1.000	0.080	0.460	0.370	1.000	0.000	0.170	1.000	1.000	0.060	0.090	0.140	1.000
MC	0.060	0.000	0.070	0.600	0.110	0.180	0.980	0.090	0.090	0.000	0.110	0.840	0.090	0.520	0.280	0.140	0.170	0.000	0.270	0.420	0.210	0.860	0.160	0.080
TO	1.000	0.070	0.000	1.000	1.000	0.990	0.040	1.000	1.000	0.110	0.000	1.000	1.000	0.340	0.210	1.000	1.000	0.270	0.000	1.000	1.000	0.210	1.000	1.000
TD	1.000	0.600	1.000	0.000	1.000	0.770	0.240	1.000	1.000	0.840	1.000	0.000	1.000	0.910	0.200	1.000	1.000	0.420	1.000	0.000	1.000	0.130	0.180	1.000
PC	0.030	0.110	1.000	1.000	0.000	0.030	1.000	0.100	0.080	0.090	1.000	1.000	0.000	1.000	0.270	0.200	0.060	0.210	1.000	1.000	0.000	1.000	0.380	0.160
RoT	0.140	0.180	0.990	0.770	0.030	0.000	1.000	0.140	0.460	0.520	0.340	0.910	1.000	0.000	1.000	0.190	0.090	0.860	0.210	0.130	1.000	0.000	1.000	0.500
Ropp	0.160	0.980	0.040	0.240	1.000	1.000	0.000	0.400	0.370	0.280	0.210	0.200	0.270	1.000	0.000	0.330	0.140	0.160	1.000	0.180	0.380	1.000	0.000	0.120
R	1.000	0.090	1.000	1.000	0.100	0.140	0.400	0.000	1.000	0.140	1.000	1.000	0.200	0.190	0.330	0.000	1.000	0.080	1.000	1.000	0.160	0.500	0.120	0.000

For away matches (-1), there is strong evidence for dependencies between Toss Decision (TD) and both Toss Outcome (TO) (1.000) and Result (R) (1.000), emphasizing their critical role in away games. In contrast, Pitch Conditions (PC) shows weaker connections with other variables (e.g., BO-PC = 0.030). In neutral venue (0) conditions, relationships such as Toss Decision (TD) and Result (R) (1.000) persist strongly, while variables like Batting Order (BO) exhibit moderate connections across factors. For home matches (1), dependencies become more selective, with fewer strong connections (e.g., PC-R = 0.160 and BO-R = 0.140), indicating a simplified network structure.

The comparative analysis across venue conditions highlights that away matches exhibit the most interconnected relationships (denser dependencies), neutral venues show intermediate dependency levels, and home matches have fewer and more streamlined dependencies. These findings suggest that external factors like venue conditions significantly influence the probabilistic relationships among variables, with away games requiring greater adaptation due to complex interdependencies.

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 2 No. 4 (November) (2024)

Network Plots

Away Venue

The network plot for the away venue in Figure 1 highlights the probabilistic relationships between key variables influencing T20I cricket outcomes. Strong dependencies are observed between Toss Outcome (TO), Toss Decision (TD), and Result (R), underscoring the importance of toss-related decisions in home games. The Ranking of Opponent (Ropp) also exhibits a significant connection with Result (R), emphasizing the impact of opposition strength even on familiar grounds. Pitch Conditions (PC) show moderate influence, linking both to Result (R) and Toss Decision (TD), suggesting that teams utilize pitch familiarity to strategize effectively.

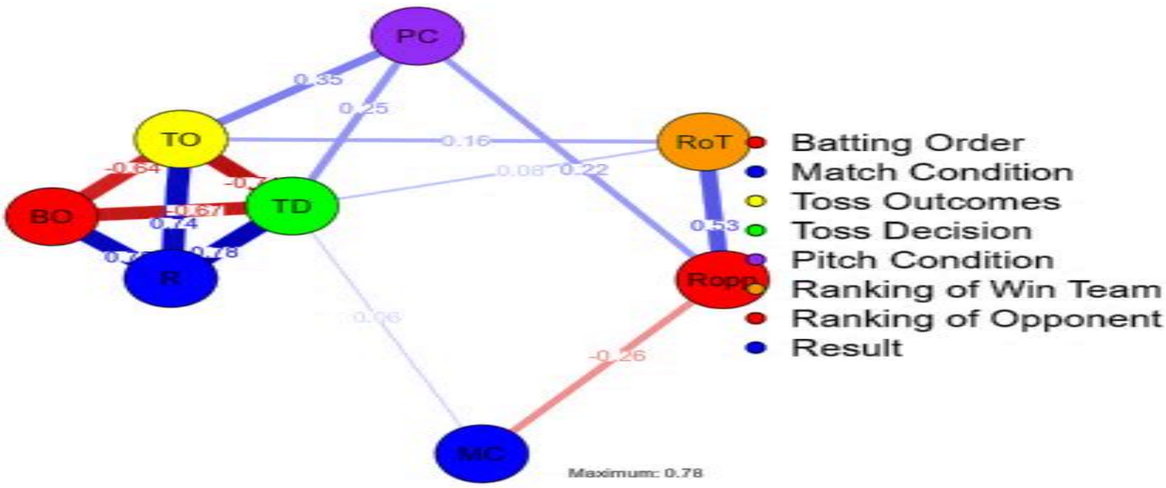

Figure 1: Network Plot for factors affecting high run chase at away venue

Batting Order (BO) moderately connects with both Result (R) and Toss Outcomes (TO), reflecting the importance of team composition in home matches. Weaker relationships, such as those involving Match Conditions (MC), indicate that external factors like day/night settings have a minimal effect on home-ground performance. Overall, the network structure demonstrates a relatively simplified dependency pattern at home; with teams leveraging controllable factors like toss decisions and pitches familiarity to optimize outcomes, while opposition strength remains a critical determinant of success.

Neutral Venue

The network plot for the neutral venue in Figure 2 depicts a more balanced dependency structure among variables affecting T20I cricket outcomes. A strong relationship persists between Toss Outcome (TO) and Toss Decision (TD), highlighting the strategic importance of toss decisions in neutral conditions. Toss Decision (TD) also connects significantly with Result (R), emphasizing its role in determining match outcomes. Pitch Conditions (PC) show notable connections with both Toss Decision (TD) and Result (R), indicating that pitch characteristics play a crucial role in neutral venues where teams lack a home-ground advantage. The Ranking of Opponent (Ropp) has a moderate link to Result (R), reflecting the influence of opposition strength.

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

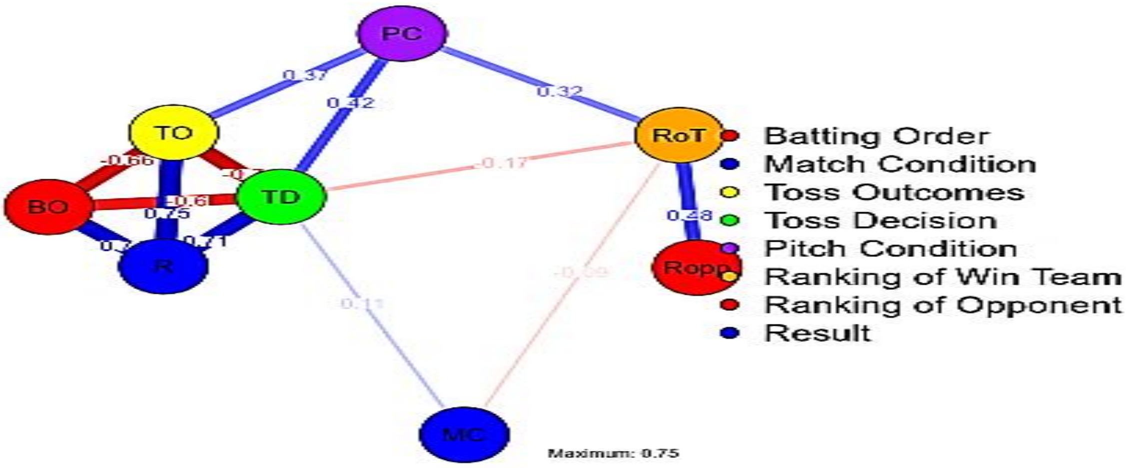

Figure 2: Network Plot for factors affecting high run chase at neutral venue

Meanwhile, Batting Order (BO) exhibits weaker connections, suggesting its reduced impact compared to other factors. Similarly, Match Conditions (MC) plays a minimal role in influencing outcomes. Overall, the neutral venue network reflects a scenario where strategy, especially around toss decisions and adaptability to pitch conditions, becomes pivotal, as neither team benefits from familiarity with the playing environment.

Home Venue

The network plot for the home venue in Figure 3 highlights the relationships between factors influencing T20I cricket outcomes when teams play at an away venue. Key variables such as Batting Order (BO), Match Condition (MC), Toss Outcomes (TO), Toss Decision (TD), Pitch Condition (PC), team rankings (RoT and Ropp), and Result (R) are interconnected.

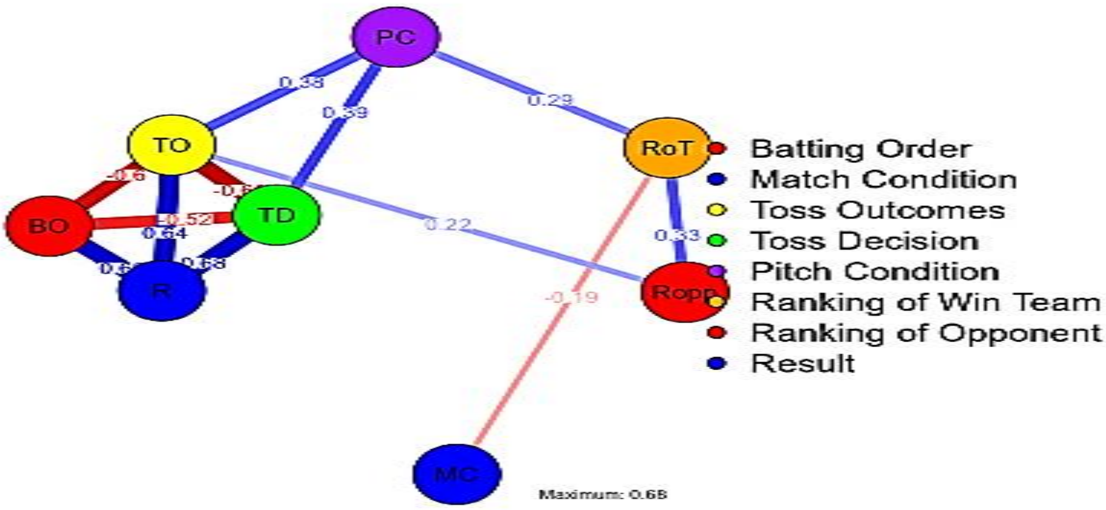

Figure 3: Network Plot for factors affecting high run chase at home venue

The strength of these relationships is represented by the thickness and color of the edges, with stronger connections like BO-TO. The plot indicates that away

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 4 (November) (2024)

venue outcomes are influenced by a combination of toss-related decisions, pitch conditions, and team rankings, showcasing the complexity of dependencies in away matches. The crossed lines in the plot are purely visual artifacts and do not impact the interpretation of the relationships.

Posterior Probability Structure Plots

The posterior probability plot for the away venue in Figure 4 depicts the posterior structure probabilities for various models, ranked by their structure index. The y-axis represents the posterior structure probability, while the x-axis represents the structure index. The plot shows a steep decline in posterior probabilities, with a few structures having relatively higher probabilities (above 0.02) and the majority having near-zero values.

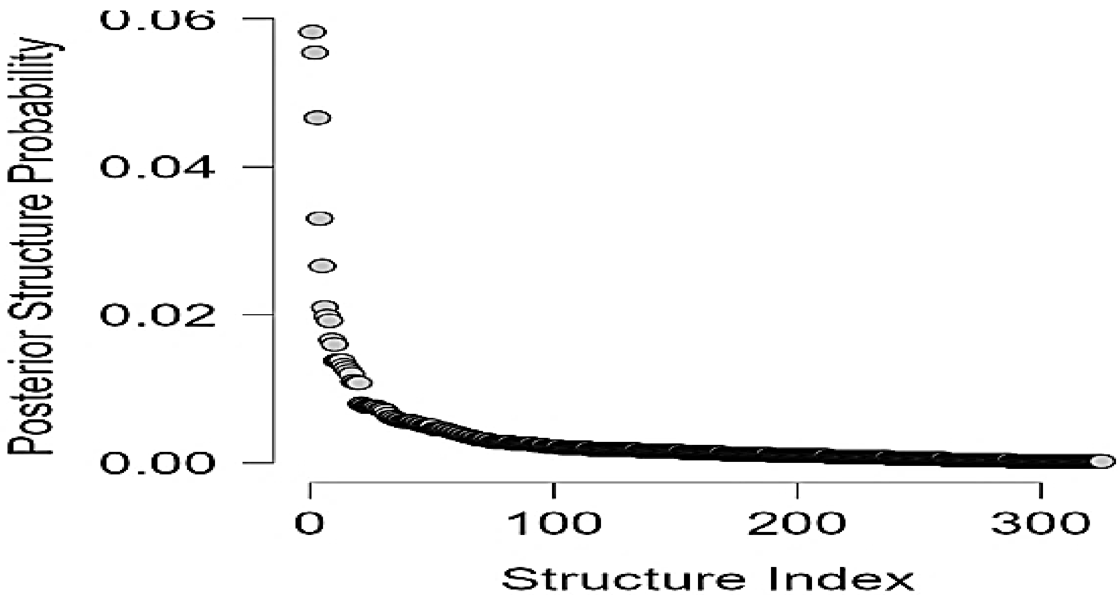

Figure 4: Posterior probability structure plot for away venue

This suggests that only a small subset of models significantly contribute to explaining the variability in outcomes for away venues, while most structures are less influential. The diminishing probabilities across the structure index highlight the dominance of a few key model structures in capturing the underlying patterns at away venues.

The posterior probability structure plot for neutral venue in Figure 5 illustrates the distribution of posterior probabilities across various structure indices. The y-axis represents the posterior structure probability, while the x-axis denotes the structure index. The plot reveals that only a small number of structures exhibit relatively higher probabilities, with the largest values concentrated near the lower structure indices (e.g., below index 50).

As the structure index increases, the probabilities rapidly decline and approach zero, indicating that the majority of structures contribute minimally to explaining the data. This distribution highlights that a few key models are highly influential in capturing the patterns, while the rest have negligible impact.

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

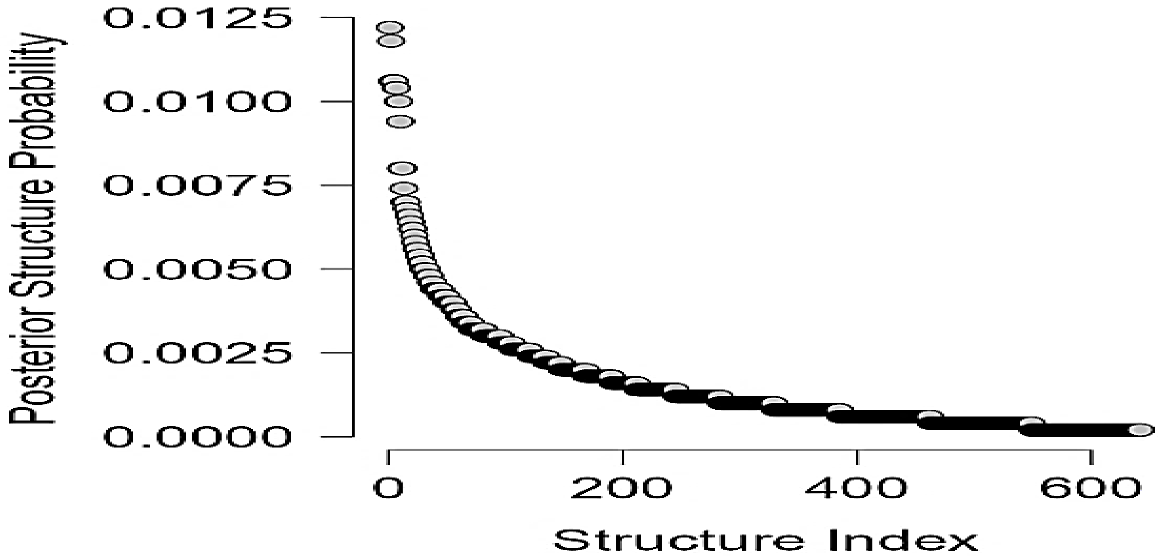

Figure 5: Posterior probability structure plot for neutral venue

The sharp decay emphasizes the sparsity of significant structures in the overall analysis.

The posterior structure probability plot in Figure 6 illustrates the probability distribution of structural indices for a given model or dataset. The x-axis represents the "Structure Index," likely indicating different model structures or configurations, while the y-axis shows the corresponding "Posterior Structure Probability," which reflects the likelihood or credibility of each structure given the data. The curve starts with relatively higher probabilities for lower indices, then sharply decreases and flattens out as the index increases, indicating that a small number of structures dominate the posterior probability distribution.

This suggests that only a few structures are strongly supported by the data, while the rest have negligible probabilities, pointing to a concentration of posterior evidence on simpler or fewer models.

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

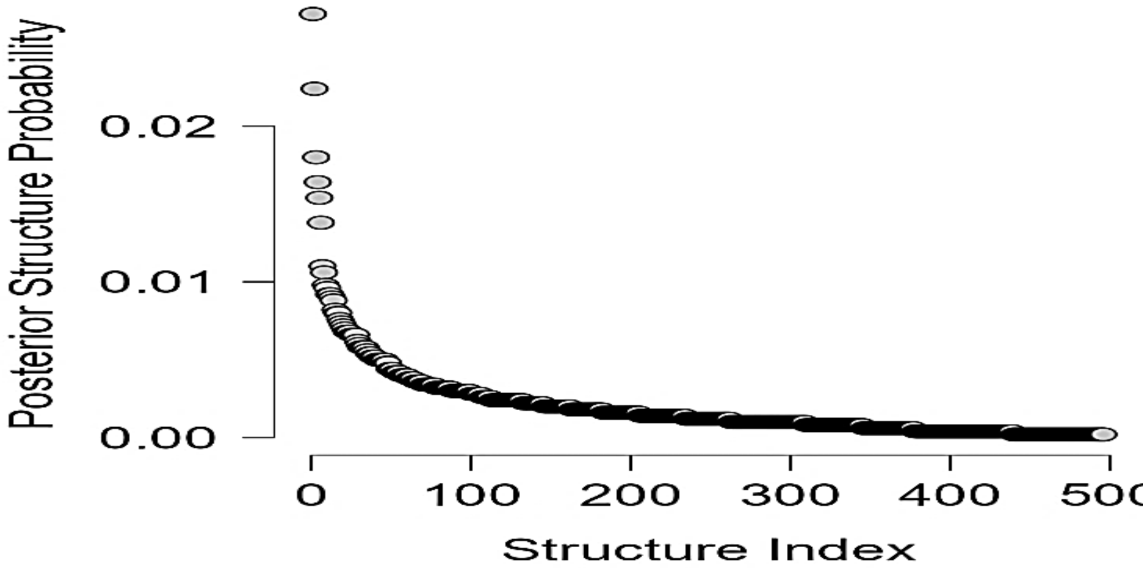

Figure 6: Posterior probability structure plot for home venue

This behavior is consistent with a Bayesian framework that penalizes complexity while favoring parsimonious models.

Table 4 summarizes centrality measures for each variable in a network analysis across three levels (-1, 0, 1), including Betweenness, Closeness, Strength, and Expected Influence. Betweenness highlights a variable's role as an intermediary in the network, with "TO" (Toss Outcome) and "PC" (Pitch Condition) showing notable bridging importance, especially at level -1. Closeness, which reflects the accessibility of a variable to others, is high for "MC" (Match Condition) and "PC" in categories 0 and 1, indicating their prominence in spreading influence. Strength, measuring the direct impact of variables, is consistently high for "R" (Result) across all levels, emphasizing its central role in the network. Finally, Expected Influence, which considers both positive and negative connections, further underscores "R" as the most influential variable, with "TO" and "PC" also exhibiting strong influences at certain levels. This analysis identifies "R," "PC," and "TO" as the most critical variables shaping the network dynamics.

Table 4: Venue wise centrality measure for factor affecting high run chase

Centrality measures per variable ▼

	-1				0				1			
Variable	Betweenness	Closeness	Strength	Expected influence	Betweenness	Closeness	Strength	Expected influence	Betweenness	Closeness	Strength	Expected influence
R	-0.719	0.374	0.808	2.063	-0.702	0.409	0.710	1.980	-0.856	0.211	0.725	1.965
BO	-0.719	0.188	0.559	-1.060	-0.702	0.245	0.546	-1.039	-0.856	0.051	0.427	-1.073
MC	-0.719	-2.050	-1.351	-0.692	-0.702	-1.784	-1.467	-0.492	-0.856	-2.108	-1.405	-0.780
TO	1.437	1.223	1.135	-0.537	-0.702	0.814	0.994	-0.760	0.856	1.102	1.301	-0.563
TD	-0.719	0.419	1.092	-0.692	1.637	1.132	1.224	-0.704	0.000	0.806	1.010	-0.610
PC	1.198	0.742	-0.816	0.467	1.169	0.604	-0.481	0.808	1.426	0.625	-0.370	0.809
RoT	-0.719	-0.443	-0.860	0.390	0.702	-0.339	-0.476	0.176	1.141	-0.247	-0.694	0.072
Ropp	0.958	-0.453	-0.568	0.061	-0.702	-1.081	-1.050	0.032	-0.856	-0.441	-0.994	0.180

Dialogue Social Science Review (DSSR)
www.journalforeducationalresearch.online

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

Centrality Plot

The centrality plot in Figure 7 depicts the relative importance and influence of various factors (Ropp, RoT, PC, TD, TO, MC, BO, R) across four centrality measures: Closeness, Betweenness, Strength, and Expected Influence, stratified by three venue levels (Vn: -1, 0, 1). Each measure captures a different aspect of factor connectivity or influence within the system. For instance, Closeness shows that certain factors (e.g., Ropp, RoT) have higher accessibility or direct influence across venue levels, with clear variations in magnitude. Betweenness centrality reveals differences in the mediating roles of factors, with some showing notable separation among venue levels, suggesting their varying ability to bridge other factors.

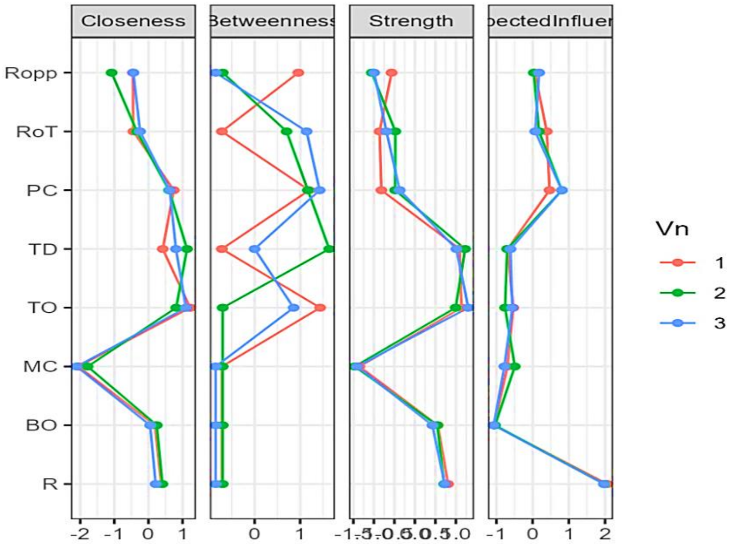

Figure 7: Venue wise centrality plot for high run chase in T20I cricket

Strength centrality indicates the cumulative influence of factors, with distinct trends observed for venue levels, emphasizing how their impact differs. Expected Influence aligns with these patterns, showcasing the overall effect of each factor on the network. The plot highlights the dynamic role of factors depending on the venue, emphasizing the complexity of their interactions.

Complexity plots

The complexity plot in Figure 8 illustrates the relationship between the number of edges in a network and the corresponding posterior probabilities for an "away" venue. The x-axis represents the number of edges (a measure of network complexity), while the y-axis shows the posterior probability of these configurations. The plot reveals that the posterior probability increases with the number of edges initially, peaking at around 15 or 16 edges, and then declines as the network becomes more complex (17 or 18 edges). This suggests that moderately complex network structures (with 15–16 edges) are most supported by the data, likely balancing model complexity and fit.

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 2 No. 4 (November) (2024)

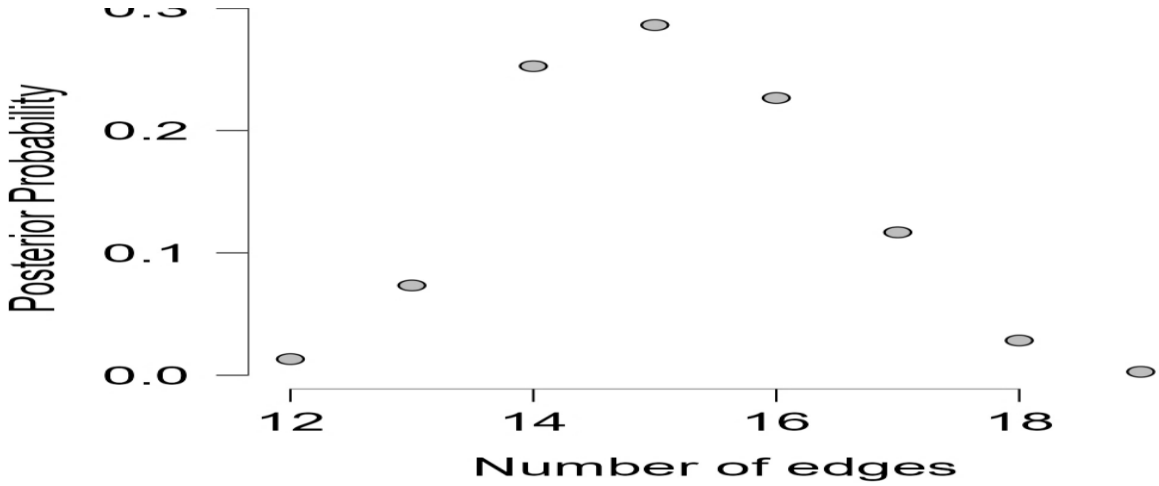

Figure 8: Away venue Complexity plot posterior probability

The drop in probability for higher edge counts indicates diminishing returns in explanatory power as the network becomes overly complex, possibly reflecting over fitting. This pattern underscores the importance of parsimony in network modeling for the "away" venue.
The complexity plot in Figure 9 for neutral venue shows the relationship between the number of edges in the model and its posterior probability for a neutral venue. The posterior probability increases as the number of edges rises from 12 to 16, indicating that adding complexity improves the model's fit up to this point.

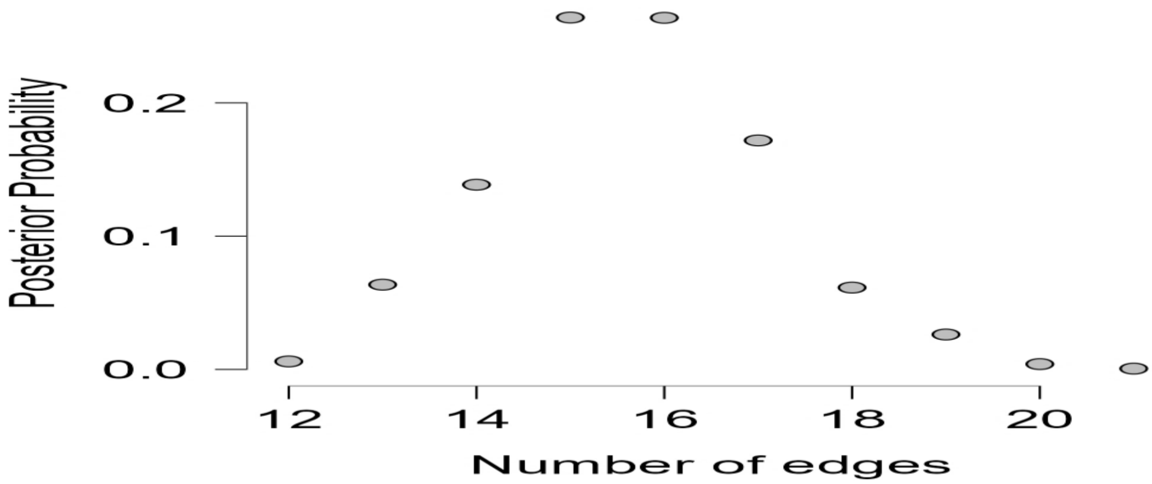

Figure 9: Neutral venue complexity plot for posterior probability

However, beyond 16 edges, the posterior probability decreases, suggesting over fitting as the model becomes overly complex. The peak at 16 edges indicates the optimal model complexity, balancing fit and simplicity, making it the best choice for modeling neutral venue conditions effectively.
The complexity plot in Figure 10 for home venue illustrates the relationship between the number of edges in the model and its posterior probability for a home venue. The posterior probability rises steadily from 12 to 16 edges, indicating that increased model complexity enhances its fit to the data up to this point. However, beyond 16 edges, the posterior probability declines, suggesting

ISSN Online: 3007-3154
ISSN Print: 3007-3146

that additional complexity leads to over fitting and reduces the model's effectiveness. The peak posterior probability at 16 edges highlights this as the optimal model complexity for capturing the key relationships while maintaining generalizability for home venue scenarios.

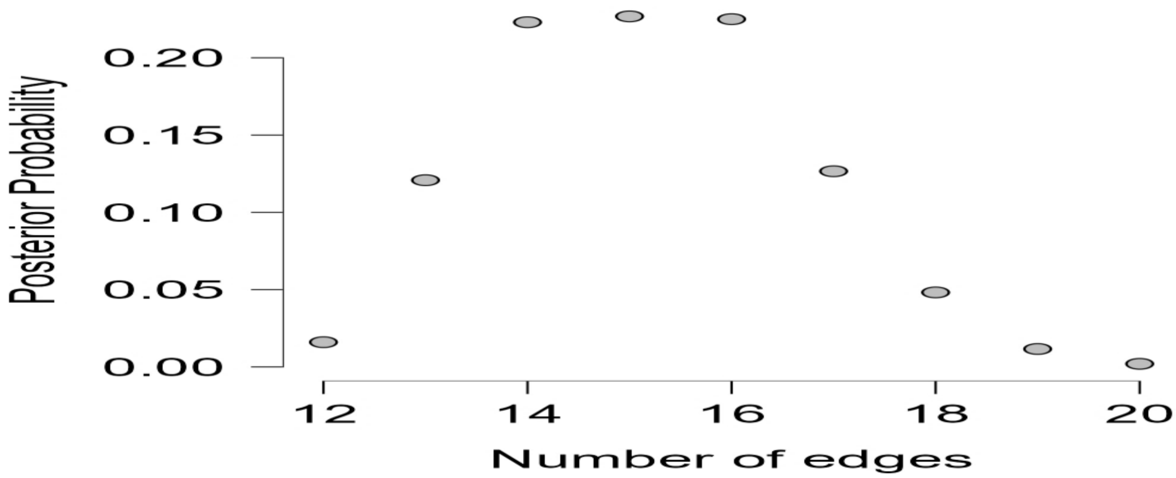

Figure 10: Home venue Complexity plot for posterior probability

Conclusion

This study analyzed the factors affecting high run chases in T20I cricket through Bayesian Networks (BNs) constructed using Gaussian Graphical Models (GGMs) across home, neutral, and away venues. The results highlighted the significant influence of venue conditions on probabilistic dependencies among variables, with away matches displaying the most complex interdependencies and home matches the simplest, reflecting varying strategic requirements. Toss Outcome (TO), Toss Decision (TD), and Pitch Conditions (PC) emerged as critical factors, with Result (R) consistently identified as the central determinant of match outcomes.

The analysis established that networks with 16 edges provided optimal complexity, achieving a balance between model accuracy and interpretability. Venue-specific differences were evident, with denser dependencies at away venues demanding greater adaptability and neutral venues emphasizing balanced strategies. Sparsity ensured the models remained interpretable, focusing on the most impactful relationships. Posterior probabilities validated the reliability of the models, confirming their suitability for predictive and strategic insights.

These findings underscore the importance of tailoring strategies to venue-specific dynamics in T20 cricket, particularly for optimizing high run chases. The integration of GGM-based BNs provides a robust, interpretable framework for cricket analytics, offering actionable insights for teams and decision-makers. Future research should expand this approach to include player-specific metrics, weather conditions, and other contextual factors to further refine predictions and strategies across different cricket formats.

References

Ahmed, R., Zaman, S., & Anwar, S. (2022). The impact of pitch conditions on batting performance in T20 cricket. *Sports Analytics Review, 15*(1), 34–47.

Clarke, S. R., & Norman, J. M. (2009). Home advantage in T20 cricket. *Journal of Sports Science and Medicine, 8*(2), 171–178.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics, 9*(3), 432–441.

Heckerman, D., & Geiger, D. (1995). Learning Bayesian Networks: A unification for discrete and Gaussian domains. *Machine Learning, 20*(3), 197–243.

Isaacs, N., & Finch, C. F. (2021). Cricket analytics: A critical review of current methods and applications. *Journal of Sports Analytics, 7*(2), 123–135.

Lauritzen, S. L. (1996). *Graphical models*. Oxford: Clarendon Press.

López-Pintado, S., González-Rubio, J., & Torres-López, T. (2021). Hybrid probabilistic models for performance prediction in cricket. *Journal of Sports Statistics, 10*(4), 121–135.

Mohan, K., Pearl, J., & Tian, J. (2013). Graphical models for probabilistic and causal reasoning. *Artificial Intelligence, 29*(6), 415–430.

Narayanan, V., & Rajagopal, S. (2020). Statistical approaches to modeling T20 cricket outcomes. *Sports Sciences Review, 28*(3), 45–60.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco: Morgan Kaufmann.

Pereira, A., Kaur, G., & Singh, R. (2015). Bayesian classifiers in sports analytics: Applications to cricket outcomes. *International Journal of Data Analysis, 12*(2), 67–82.

Peters, J., Bühlmann, P., & Meinshausen, N. (2014). Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(5), 947–1012.

Pollard, R. (2008). Home advantage in cricket: International and domestic results. *International Journal of Performance Analysis in Sport, 8*(1), 49–63.

Samad, M. E. (2019). Probabilistic approaches for identifying run-scoring advantages in cricket. *Journal of Sports Analytics, 5*(1), 65–79.

Shah, S. A. A., Zaman, Q., Hussain, S., Iftikhar, S., Shah, S. H., Sabir, A. R., & Shah, A. (2024). Analyzing the impact of venue and match factors on ground-wise average scores and successful run chases in T-20I cricket using factorial design. Kurdish Studies, 12(5), 781-788.

Smith, T., Wilson, J., & Miller, K. (2021). Incorporating domain knowledge in probabilistic models for cricket analytics. *Sports Analytics Research, 9*(2), 89–101.