

Article

Not peer-reviewed version

Sports Analytics with Graph Neural Networks and Graph Convolutional Networks

[Tim Todd Drexler](#)*

Posted Date: 1 October 2024

doi: 10.20944/preprints202410.0046.v1

Keywords: Graph Neural Networks; Graph Convolutional Networks



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Sports Analytics with Graph Neural Networks and Graph Convolutional Networks

Tim Todd Drexler

Affiliation 1; timtodddrexler@outlook.com

Abstract: Graph neural networks and graph convolutional networks are effective at representing complex relationships. We attempt to apply them to sports analytics by reviewing existing applications in sports. These methodologies leverage the inherent graph structure of sports data, capturing interactions among players, teams, and game events. By systematically examining various studies, we highlight how GNNs and GCNs can enhance performance prediction, player evaluation, and tactical analysis in sports. Our review emphasizes the unique advantages these models offer over traditional analytical approaches, particularly in their ability to account for the dynamic and interconnected nature of sports activities. We conclude by identifying potential areas for future research and application, suggesting that the integration of advanced graph-based techniques could significantly advance the field of sports analytics.

Keywords: Graph Neural Networks; Graph Convolutional Networks

1. Introduction

Sports starts to leverage on machine learning in the new century. The integration of data-driven methods in sports analytics has evolved significantly with the advent of machine learning techniques, making it possible to extract meaningful insights from complex and large datasets. In particular, Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs) have garnered increasing attention due to their ability to model structured data, such as player interactions, team formations, and game dynamics, as graphs. Unlike traditional machine learning methods, GNNs and GCNs capture spatial and temporal dependencies between entities, providing deeper insights into strategic elements in sports. This paper delves into the application of GNNs and GCNs in sports analytics, exploring their potential to enhance performance prediction, player scouting, tactical analysis, and injury prediction.

2. Review of Literature

2.1. Overview of Graph-Based Learning

Graphs are powerful data structures that can model relationships between entities. In a typical graph $G = (V, E)$, where V represents nodes (e.g., players or teams), while E represents edges (e.g., passes between players or games between teams). Graph-based learning leverages these structures to analyze the intricate relationships that exist in various domains, including sports.

GNNs and GCNs, as subclasses of deep learning models, extend this capability by applying neural network techniques directly to graph-structured data. GNNs use recursive neighborhood aggregation to capture the influence of a node's neighborhood on its representation. GCNs, on the other hand, focus on convolution operations that capture localized graph features by combining node features with their neighboring nodes.

2.2. GNNs in Sports

Graph Neural Networks (GNNs) have been gaining traction in sports analytics due to their capability to model complex relationships between players, teams, and game events. Sports are inherently relational; they involve interactions between players, coordination within teams, and spatial-temporal dynamics, all of which can be effectively represented using graph-based models.

GNNs provide a powerful way to analyze this data by capturing both local and global dependencies within the game structure. This section reviews the existing literature that applies GNNs in various sports analytics tasks.

2.2.1. Tactical Analysis and Team Performance

Tactical analysis in team sports such as soccer, basketball, and American football is an area where GNNs have shown great promise. GNNs allow for the modeling of player interactions and their movements on the field as a dynamic graph, where players are nodes and edges represent their interactions (e.g., passes, tackles, or assists).

- Soccer: In soccer, researchers have used GNNs to model formations and passing networks to predict the success of a play or identify tactical weaknesses in a team. For example, the work by [1] explores the use of GNNs for understanding soccer formations and player positioning in real-time. Their model aggregates spatial information from neighboring players to analyze the influence of different formations on game outcomes. [2] emphasized on formation selection. This kind of application shows how GNNs can be used to evaluate and optimize team strategies in real time.
- Basketball: GNNs have also been applied in basketball to predict player movements and decision-making during offensive and defensive plays. [3] developed a GNN-based model to predict the effectiveness of player actions based on their spatial positioning and interactions with opponents. Their model was able to predict the success of passes, shots, and other events based on player movement graphs, significantly improving prediction accuracy compared to traditional machine learning approaches.

2.2.2. Player Performance and Injury Prediction

Another area of research in sports analytics with GNNs is player performance modeling and injury prediction. GNNs can model the interaction between players and their environment, including the influence of teammates, opponents, and the physical strain of the game.

- Player Performance: GNNs have been applied to model individual player performance by incorporating both spatial and temporal information. [4] proposed a GNN-based framework to predict player performance in soccer by analyzing historical game data. Their model accounts for player interactions within the team as well as game context (e.g., score, time, and possession). This allows for a holistic prediction of player contributions to team success.
- Injury Prediction: Injuries in sports often result from complex interactions between players, their movements, and external factors like game intensity. [5] used GNNs to predict injury risk by modeling the physical strain experienced by players based on their movements and interactions. Their model aggregates data from wearable sensors and game statistics, transforming it into a graph structure to predict injury risks more accurately than previous statistical methods.

2.2.3. Game Outcome Prediction

Game outcome prediction, which involves predicting the final result of a match based on player and team performance, has also benefited from the application of GNNs. GNNs can model how different teams and players interact, and use this information to predict game outcomes.

- Football: [6] applied GNNs to model American football games by representing players and their interactions as nodes and edges in a graph. Their model captured both individual player actions and team-level strategies, providing a significant improvement in outcome prediction compared to traditional logistic regression and random forest models.
- eSports: In the field of eSports, GNNs have been used to predict the outcomes of competitive video games. [7] applied GNNs to analyze player interactions in games such as Dota 2 and League of Legends. Their model considers both player performance and team coordination,

providing more accurate predictions compared to traditional game outcome prediction models. We are not going to put our main focus on the eSports for the current study.

2.2.4. Challenges in GNNs for Sports Analytics

Despite the growing use of GNNs in sports, there are several challenges that remain. One of the biggest challenges is the availability of high-quality data. Sports data is often proprietary, meaning that research in this field is often limited to professional teams with access to sophisticated tracking systems. Moreover, GNN models can be computationally expensive, especially when dealing with dynamic and large-scale data.

Additionally, GNN models in sports require significant domain expertise to construct the graph in a meaningful way. The structure of the graph (nodes and edges) can vary between sports, and the choice of features included in the model has a large impact on its performance. Finally, interpretability remains a challenge, as GNNs are often considered "black box" models, which can limit their usability in real-world sports decision-making.

2.3. GCNs in Sports

While Graph Neural Networks (GNNs) encompass various architectures designed for graph-structured data, Graph Convolutional Networks (GCNs) are specifically designed to apply convolutional operations on graphs. They provide a more structured and efficient approach for learning representations of nodes (e.g., players) in graph-based sports analytics, leveraging localized graph convolution to capture spatial and relational information.

In sports, GCNs have been particularly effective in tasks such as performance evaluation, tactical analysis, and outcome prediction, where team dynamics and player interactions can be modeled as graphs. Unlike generic GNN models, GCNs use convolution-like operations on graph nodes, allowing the model to aggregate features from neighbors, which makes them particularly suitable for applications requiring spatial awareness and interdependence modeling.

2.3.1. GCNs for Tactical and Spatial Analysis

In team sports like soccer, basketball, and football, tactical analysis is vital for understanding team formations, player movements, and the effectiveness of strategic setups. GCNs have been pivotal in capturing these spatial relationships. While GNNs can model general interactions, GCNs excel in explicitly learning the spatial features of players (nodes) within the graph.

Soccer Tactical Setup

- Soccer Tactical Setup and Passing Networks: In soccer, passing sequences and positional dynamics between players are fundamental to understanding team strategies. GCNs have been utilized to build passing networks, where the players are the nodes, and passes between them form the edges. A key difference in GCNs is that they allow not only an understanding of passing frequency but also of the "importance" or influence of certain nodes (players) within a tactical setup.
- [1] applied GCNs to capture a team's overall structure and predict passing outcomes based on positional play. The GCN aggregated local features of players' movements and passing actions, and its convolutional layers allowed the model to account for the spatial relationships critical to team performance.

2.3.2. Spatial Awareness in Basketball

Player Movements and Court Positioning: In basketball, spatial relationships between players (defensive setups, offensive screens, etc.) are paramount. GCNs are uniquely positioned to handle this kind of spatial data due to their ability to capture both local (one player's movement relative to their immediate neighbors) and global (overall team formation) structures.[8] employed GCNs to analyze on-court player movement and interactions. The network processed the positions of all players at any

given moment to analyze the evolving offensive or defensive formations. The convolutional layers in the GCN captured the spatial-temporal evolution of these formations, highlighting how players' proximity and movement patterns affected overall team strategy.

2.3.3. Player Interaction and Performance Evaluation

Another key application of GCNs in sports is their use in evaluating individual and team performance through player interactions. Unlike basic GNN models, GCNs are more capable of capturing player dependencies while also recognizing the spatial context in which these interactions occur.

Soccer Player Networks

Evaluating Contributions Beyond Ball Possession: In soccer, GCNs have been utilized to understand player contributions that go beyond simple metrics like ball possession or goals. By building player interaction graphs, where nodes are players and edges represent passes or other interactions, GCNs can evaluate how a player's position and movements affect team performance. [9] applied GCNs to build player performance models that incorporated both individual statistics and interaction-based metrics. The model aggregated information from neighboring players (via graph convolutions), allowing it to evaluate a player's influence on overall team dynamics, especially in off-the-ball movements where traditional metrics fail.

Tennis and Sequential Event Graph

Individual sports like tennis, while not relying on traditional team-based tactics, still benefit from GCNs when analyzing player performance through match sequences. By representing match events (serves, volleys, returns) as graph nodes and the temporal sequence of these events as edges, GCNs can provide insights into the effectiveness of a player's strategic decisions. [10] used GCNs to analyze event sequences in tennis. They demonstrated how GCNs, by applying convolution over event sequences, could capture critical decision points in matches. For instance, a player's decision to approach the net at certain points in a rally could be understood not in isolation but in relation to earlier moves and positional dynamics modeled through the graph.

2.3.4. Game Outcome and Play Prediction

GCNs have been extensively used in predicting game outcomes, with applications that emphasize their ability to model the interdependencies between player actions and overall team performance. Unlike GNNs, which can work on dynamic or evolving data, GCNs excel when game states or play-by-play actions can be structured as static graphs, allowing convolutional layers to capture both local and global game states.

- **Football Play Prediction:** American Football and Play Outcome Prediction: In American football, each play can be modeled as a static graph where players are nodes, and interactions (e.g., blocks, tackles) form edges. GCNs are particularly effective at handling this kind of data because they can capture how individual player actions influence the overall outcome of a play. [6] applied GCNs to predict the success of football plays. By structuring each play as a graph, with nodes representing players and edges representing player actions, the GCN aggregated information across the graph to predict whether a play would result in a successful pass or run. The convolutional layers allowed the model to consider not just immediate interactions but the overall configuration of players on the field.
- **Predicting Outcomes in eSports:** Teamfight Tactics in eSports: GCNs have also been applied to eSports, where team-based strategy games like Dota 2 or League of Legends involve complex player interactions. GCNs can effectively model player decisions, actions, and interactions within a single game state. [11] used GCNs to predict game outcomes in eSports by structuring the game's player interactions as a graph. The convolutional layers allowed the model to learn from

a player's immediate neighbors and aggregate higher-level information about the team's overall strategy, providing accurate predictions of game outcomes based on teamfight dynamics.

2.3.5. GCN-Specific Challenges in Sports

While GCNs have shown great promise, they also present certain challenges specific to sports applications:

- **Spatial Data Sparsity:** GCNs rely heavily on the quality and quantity of spatial and interaction data. In many cases, the data required to build detailed graphs for sports analysis (such as player tracking data) is sparse or difficult to obtain. Even when data is available, building accurate and meaningful graphs that capture the full complexity of player interactions can be a challenge.
- **Real-Time Processing and Scalability** GCNs, while efficient in handling spatial data, are computationally intensive when applied to large, dynamic sports datasets. For real-time game analytics, where decisions need to be made quickly (e.g., in-game tactical adjustments), the computational overhead of GCNs can be a bottleneck. Future research will need to address the trade-off between model accuracy and computational efficiency, particularly in real-time sports analytics.

3. Methodologies for Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs) in Sports Analytics

3.1. 1. Methodology for Graph Neural Networks (GNNs) in Sports Analytics

GNNs are powerful tools for processing data represented as graphs. In sports analytics, they can be used to model relationships among players, teams, and game events. The following steps outline a general methodology for applying GNNs in sports analytics:

First, **Graph Construction** involves defining the nodes (e.g., players, teams, matches) and edges (e.g., interactions such as passes, tackles, shots). A graph structure is created where the adjacency matrix represents the connections between nodes, and feature vectors describe node characteristics (e.g., player stats, position, historical performance). Next, **Feature Engineering** entails developing relevant features for each node and edge. For instance, player features may include metrics like goals, assists, and distance covered, while edge features may represent interaction types (e.g., successful passes, assists). It is essential to normalize and preprocess features for improved model performance.

The next step is to choose an appropriate **Model Architecture**, such as Message Passing Neural Networks or Graph Attention Networks. Implement layers that allow nodes to aggregate information from their neighbors iteratively. In the **Training** phase, use labeled data for supervised learning (e.g., predicting match outcomes or player performance). Define loss functions relevant to the task (e.g., cross-entropy for classification tasks, mean squared error for regression tasks), and optimize the model using an optimizer like Adam or SGD.

After training, **Evaluation** requires splitting the dataset into training, validation, and test sets. Use metrics like accuracy, precision, recall, and F1 score to evaluate model performance, and perform ablation studies to assess the impact of different features or architectural choices. Finally, **Application** of the trained GNN model can include real-world scenarios such as predicting player transfers, optimizing team formations, or identifying potential injury risks based on player interactions.

An example of GNN application is Player Interaction Analysis in Soccer. In this case, graph construction involves creating a graph where nodes are players in a soccer match and edges represent successful passes between them. Feature engineering entails assigning features such as player position, total passes, assists, and passing accuracy. For model architecture, a Graph Attention Network (GAT) can be used to allow the model to focus on important interactions between players. During training, the model is trained to predict the likelihood of a successful pass based on player positions and historical interaction data. The evaluation can be conducted using metrics like accuracy and F1 score, comparing the GNN's predictions against traditional models that consider players in isolation.

3.2. 2. Methodology for Graph Convolutional Networks (GCNs) in Sports Analytics

GCNs are specifically designed for performing convolution operations on graph-structured data. The following methodology outlines the steps for applying GCNs in sports analytics. First, **Graph Representation** involves defining the graph structure with nodes representing entities (e.g., players, games) and edges representing relationships (e.g., passes, fouls). An adjacency matrix is used to represent the connections between nodes.

Next, **Feature Preparation** requires collecting and engineering features for each node and edge. For example, in soccer, player features may include distance covered, goals scored, and position on the field. Edge features could represent interaction types (e.g., successful passes, defensive actions). Following this, **GCN Layer Design** involves implementing GCN layers that perform convolution over the graph structure, allowing nodes to aggregate information from their neighbors. The aggregation can be weighted based on edge features or use methods like attention mechanisms.

During the **Training** phase, a labeled dataset is used to train the GCN for supervised learning tasks (e.g., classifying player performance). Loss functions appropriate for the task are defined, and optimizers are used for model training. Validation and testing require splitting data into training and test sets. The model's performance is evaluated using appropriate metrics, and cross-validation ensures robustness. Finally, the **Deployment** of the trained GCN model allows for analyzing ongoing sports events, evaluating player performance in real-time, or simulating different game strategies.

An example of GCN application is Game Outcome Prediction in American Football. In this case, graph representation involves creating a graph where nodes represent players and edges represent interactions during a play (e.g., tackles, assists). Feature preparation entails assigning features like player speed, average yards gained, and positions on the field. For GCN layer design, GCN layers can be implemented to allow each player to aggregate features from their teammates and opponents, learning the influence of player positioning on play outcomes. During training, the GCN predicts the likelihood of a successful play (e.g., a touchdown) based on the interactions modeled in the graph. Validation and testing involve evaluating model performance on historical game data using metrics such as accuracy and area under the curve (AUC).

4. Conclusion

In conclusion, both GNNs and GCNs provide powerful methodologies for analyzing complex interactions and relationships in sports data. By leveraging graph structures, these models can uncover insights that traditional methods may overlook, such as the impact of player positioning and interactions on performance. Through careful design and application, GNNs and GCNs can significantly enhance sports analytics capabilities, leading to more informed decision-making in areas like player evaluation, tactical analysis, and outcome prediction.

References

1. Mao, Z.; Zhang, L.; Xu, F. Graph Neural Networks for Soccer Tactical Analysis. *Journal of Sports Analytics* **2021**, pp. 1–12.
2. Wang, Z.; Zhu, Y.; Li, Z.; Wang, Z.; Qin, H.; Liu, X. Graph neural network recommendation system for football formation. *Applied Science and Biotechnology Journal for Advanced Research* **2024**, *3*, 33–39.
3. Qi, X.; Liu, J.; Zhang, J. Predicting Basketball Player Movements with Graph Neural Networks. *Proceedings of the 2020 ACM SIGKDD Conference* **2020**, pp. 234–242.
4. Wang, M.; Zhang, W. Player Performance Prediction Using Graph Neural Networks in Soccer. *IEEE Transactions on Sports Analytics* **2022**, pp. 112–125.
5. Zhang, H.; Li, M. Injury Risk Prediction in Professional Sports Using Graph Neural Networks. *arXiv preprint arXiv:1912.00555* **2019**.
6. Li, J.; Wang, H. Graph Neural Networks for American Football Play Outcome Prediction. *Proceedings of the 2020 AAAI Conference* **2020**, pp. 3421–3429.
7. Sharma, A.; Patel, J. Using Graph Neural Networks for Game Outcome Prediction in eSports. *ACM Transactions on eSports Analytics* **2021**, pp. 88–103.

8. Li, G.; Muller, M.; Thabet, A.K.; Ghanem, B. DeepGCNs: Can GCNs Go As Deep As CNNs? *Proceedings of the IEEE/CVF International Conference on Computer Vision* **2019**, pp. 9267–9276.
9. Wang, M.; Zhang, W. Evaluating Soccer Player Performance Using Graph Convolutional Networks. *IEEE Transactions on Sports Analytics* **2021**, pp. 110–122.
10. Liu, X.; Wu, J.; Zhang, H. Using Graph Convolutional Networks for Tennis Match Analysis. *Journal of Sports Performance Analysis* **2021**, pp. 234–245.
11. Wang, L.; Zhang, F. Game Outcome Prediction in eSports Using Graph Convolutional Networks. *ACM Transactions on eSports Analytics* **2020**, pp. 78–92.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.