Go through the description of this dataset and give me some project ideas that can be implemented using ML with Graphs , Don't give the code just give the ideas and workflow

Based on my analysis of the StatsBomb Open Events dataset and research on graph-based machine learning applications in football, I can provide you with several innovative project ideas that leverage the rich relational and temporal data in this dataset.

Machine Learning with Graphs: Football Analytics Project Ideas

1. Dynamic Pass Network Analysis for Tactical Pattern Recognition

Core Idea: Build dynamic passing networks that evolve over time to identify tactical patterns and predict team strategy changes.

Workflow:

- Create temporal graphs where nodes represent players and edges represent passes with weights indicating frequency, success rate, and timing
- Use Graph Convolutional Networks (GCNs) to learn player embeddings that capture positional and tactical roles
- Apply temporal graph neural networks to model how passing patterns evolve during different game phases
- Cluster similar tactical patterns and predict when teams are likely to change formation or strategy
- Incorporate "under_pressure" and "counterpress" attributes to understand defensive influence on passing networks

Key Features from Dataset: Pass events, player positions, timestamps, formation data, pressure events, tactical shifts

2. Player Influence and Impact Prediction using Graph Centrality

Core Idea: Develop a GNN-based system to identify "hidden influencers" - players whose actions significantly impact team performance beyond traditional statistics.

Workflow:

- Construct multi-layered graphs combining passing, defensive actions, and spatial relationships
- Calculate dynamic centrality measures (betweenness, closeness, eigenvector centrality) that change during the match
- Use Graph Attention Networks (GAT) to learn which players have the most influence on possession chains and goal-scoring opportunities
- Predict player ratings that go beyond goals/assists to capture defensive contributions, buildup play, and space creation
- Validate against outcomes like "Expected Threat (xT)" changes and possession success

Key Features from Dataset: All event types, player positions, possession chains, related_events, counterpress indicators

3. Counterattack Success Prediction with Spatio-Temporal GNNs

Core Idea: Predict the likelihood of successful counterattacks using graph representations of player positions and movements.

Workflow:

- Identify counterattack sequences using possession changes and "From Counter" play patterns
- Create dynamic graphs where nodes are players with features like position, speed, distance to goal
- Edges represent spatial relationships and potential passing lanes
- Use spatio-temporal GNNs to model how counterattack opportunities evolve
- Predict success probability based on player positioning, defensive pressure, and ball progression patterns
- Incorporate "carry" events and "dribbled_past" to model individual breakthrough moments

Key Features from Dataset: Play patterns, carry events, dribble events, pressure events, shot events, location coordinates

4. Defensive Coordination and Pressing Analysis

Core Idea: Analyze defensive structures using graph neural networks to understand coordinated pressing and space control.

Workflow:

- Build defensive interaction networks where nodes are defending players and edges represent coordination patterns
- Use "counterpress" and "pressure" events to identify coordinated defensive actions
- Apply GNNs to learn defensive formation patterns and predict pressing success
- Model how defensive actions (blocks, interceptions, clearances) cascade through the defensive network
- Predict opponent attacking threat reduction based on defensive coordination quality

Key Features from Dataset: Pressure events, counterpress indicators, block events, interception events, clearance events, duel outcomes

5. Shot Context and Freeze Frame Analysis with Graph Learning

Core Idea: Use the freeze frame data to create spatial graphs at shot moments and predict shot outcomes.

Workflow:

- Extract freeze frame data for each shot to create snapshot graphs of player positions
- Nodes represent players with features like position, distance to goal, angle, team affiliation
- Edges encode spatial relationships, marking assignments, and defensive coverage
- Use Graph Neural Networks to learn optimal attacking and defensive positioning patterns
- Predict shot success probability, goalkeeper action type, and optimal defensive response
- Compare model predictions with actual "statsbomb_xg" values to validate approach

Key Features from Dataset: Shot freeze frames, goalkeeper events, shot outcomes, player positions, expected goals data

6. Formation Evolution and Tactical Adaptation Prediction

Core Idea: Track and predict how team formations evolve during matches using dynamic graph representations.

Workflow:

- Model team formations as graphs where nodes are positions and edges represent tactical relationships
- Track formation changes using "Starting XI" and "Tactical Shift" events

- Use temporal GNNs to learn how formations adapt to game state (score, time, opponent actions)
- Predict when teams are likely to make tactical changes based on performance metrics
- Identify optimal counter-formations against specific tactical setups

Key Features from Dataset: Tactical lineups, tactical shifts, formation data, match context, score progression

7. Multi-Modal Event Sequence Prediction

Core Idea: Predict the next sequence of events using graph representations that combine multiple event types.

Workflow:

- Create heterogeneous graphs with different node types (players, positions, events)
- Use possession chains to model event sequences and their relationships
- Apply Graph Transformer networks to capture long-range dependencies between events
- Predict next event type, location, and success probability
- Incorporate contextual information like game phase, score difference, and player fatigue indicators

Key Features from Dataset: Event sequences, possession chains, event types, timestamps, related events, play patterns

8. Injury Risk and Player Load Analysis through Movement Networks

Core Idea: Analyze player movement patterns and interaction intensity to predict injury risk and optimal rotation strategies.

Workflow:

- Build player interaction networks based on duels, challenges, and physical contests
- Use carry events and location data to model player movement intensity
- Apply GNNs to identify players with high interaction loads and stress patterns
- Predict injury risk based on network centrality and interaction frequency
- Suggest optimal substitution timing and rotation strategies

Key Features from Dataset: Duel events, carry events, foul events, injury stoppages, substitution events, player movements

Technical Implementation Considerations

Graph Construction:

- Use sliding time windows to create temporal snapshots
- Implement multi-edge graphs to capture different relationship types
- Apply graph sampling techniques for computational efficiency

Model Architecture:

- Combine GCNs for local feature learning with attention mechanisms for important relationship identification
- Use recurrent components for temporal modeling
- Implement multi-task learning to jointly predict multiple outcomes

Evaluation Metrics:

- Compare against baseline statistical models
- Use cross-validation with time-series splits to prevent data leakage
- Validate predictions against actual match outcomes and expert tactical analysis

These projects leverage the rich relational structure in the StatsBomb dataset while addressing real analytical needs in football. Each project can provide valuable insights for coaches, analysts, and sports scientists while advancing the application of graph neural networks in sports analytics.

⁂

1. Open-Data-Events-v4.0.0.pdf
2. https://kth.diva-portal.org/smash/get/diva2:1845172/FULLTEXT01.pdf
3. https://www.sloansportsconference.com/research-papers/detection-of-tactical-patterns-using-semi-supervised-graph-neural-networks
4. https://arxiv.org/html/2411.17450v2
5. https://arxiv.org/html/2503.09737v1
6. https://arxiv.org/html/2409.13098v1
7. https://www.statsperform.com/wp-content/uploads/2021/04/Making-Offensive-Play-Predictable.pdf
8. https://www.sciencedirect.com/science/article/pii/S2405844023095245
9. https://www.sciencedirect.com/science/article/abs/pii/S0950705121002604
10. https://arxiv.org/html/2502.15462v1
11. https://www.nature.com/articles/s41598-025-05462-7